



Increasing the Ability of an Experiment to Measure an Effect

Print

Return

Sandra Slutz, PhD
Staff Scientist, Science Buddies

Kenneth L. Hess
Founder and President, Science Buddies

All experimental observations are a combination of *signal*, the true effect of a variable on an outcome, and *noise*, the random error inherent in your experimental technique. When designing and analyzing experiments, the goal is to maximize the signal-to-noise ratio so that you can draw accurate conclusions. Six common means of increasing the signal-to-noise ratio are:

1. Making repeated measurements of one item,
2. Increasing sample size,
3. Randomizing samples,
4. Randomizing experiments,
5. Repeating experiments, and
6. Including covariates.

Rarely would any one scientific question lend itself to using all six of these signal-to-noise reducing techniques, but to maximize the accuracy of your results, you should try to incorporate as many of these as *reasonably* possible into your experimental designs. Cost, time, and resource availability will help dictate which techniques are possible and which aren't. Table 1, below, outlines the types of scenarios where each technique is most useful.

Quantitative Variables			
Technique for increasing the signal-to-noise ratio	What is it?	When is it helpful?	Examples of when to use it
Making repeated measurements	<p>Measuring a single item or event more than once to eliminate error in measuring.</p> <p>More measurements of a single event lead to greater confidence in calculating an accurate average measurement.</p>	<p>Especially helpful if an individual measurement may have a lot of variability; because it has to be made quickly, it is hard to determine the exact endpoint, or is technically difficult and thus prone to errors.</p> <p>Does not add value if the measurement is clear-cut, like the answer to a survey question about a person's age or measuring the dimensions of a room in meters.</p>	<ul style="list-style-type: none"> • How many drops of acid does it take to change the color of this indicator solution? Run the reaction several times on aliquots of the same solution. • How long does it take for this specific graphics card to heat the air surrounding it to 100°C? Test the same exact graphics card multiple times. • How long does this turtle spend underwater before surfacing for a breath? Observe the same turtle multiple times.
Increasing the sample size	<p>Increasing the number of items, or people, that you are collecting data from increases the probability that what you are observing is indicative of the whole population.</p> <p>Calculations can be made to determine how large the sample size needs to be. See the guide on determining the best sample size for a survey for more details.</p>	<p>Especially helpful when you are trying to draw conclusions about an entire population.</p> <p>Does not apply if your conclusions are intended to be specific to an individual or single item.</p>	<ul style="list-style-type: none"> • Do teenagers eat healthy foods? Survey a large number of teens, not just five people who always hang out together, about their daily diets. • How do the lung capacities of smokers versus non-smokers compare? Take measurements from many smokers and non-smokers. • How long does a 9-volt (V) battery from brand X power a flashlight? Test multiple manufacturing batches of brand X's 9-V battery.
Randomization of samples	<p>Using a lottery system to assign samples to different experimental and control groups within a given experiment helps make the starting makeup of the groups as equal as possible, even for variables you might have overlooked.</p> <p>Some experiments can be completely randomized; other involve <i>blocking</i> first. Blocking allows for the creation of homogenous groups, like males versus females, and then randomization within the block. This variation is done when the researcher suspects that there may be scientifically important differences between experimental subjects.</p>	<p>Especially critical when the population you are drawing your samples from (which is the population you want to make conclusions about) is very heterogeneous.</p> <p>May not apply if you need to stratify your population because you want to be able to draw different conclusions about each sub-group. For example, men vs. women in a drug study or different types of resistors in a circuit design.</p>	<ul style="list-style-type: none"> • Which fertilization technique increases crop yield the most? Assign fertilizer treatment to each plot of land by lottery, thus evening out effects of other variables, like soil makeup and water content, among the experimental groups. • Does this medication decrease osteoporosis? Randomly assign people to determining whether a medication is effective. Randomly assign patients to placebo or medication group.
Randomization of experiments	<p>Using a lottery system to determine the order of carrying out related experiments, rather than relying on an apparently logical order that may introduce other overlooked variables.</p>	<p>Especially critical when you have related experiments from which you are going to draw a single meta conclusion.</p> <p>Applies to both related experiments done serially using the same equipment, and related experiments done in parallel on different equipment.</p>	<ul style="list-style-type: none"> • Does the length of time plastic is pressed in a mold affect the final strength of the plastic? Rather than running experiments testing 10, 20, 30, etc. seconds of pressing back to back, randomize which length of time is tested first, second, etc. The randomization eliminates potential effects from other variables like different amounts of mixing time for the plastic as the experiments progress and changes to the temperature of the mold over the course of all the experiments. • Does the color of a mouse maze walls affect the total time it takes for mice to find their way through? For a single experimental repeat, test all the mice on the same day in all the mazes. Mazes should be identical, other than wall color. For each mouse, randomly determine in which order he should be tested in the different-colored mazes. The randomization will eliminate

			potential contribution of effects like mouse fatigue over the course of the testing.
Repeating experiments	Repeating an experiment more than once helps determine if the data was a fluke, or represents the normal case. It helps guard against jumping to conclusions without enough evidence. The number of repeats depends on many factors, including the spread of the data and the availability of resources. Three repeats is usually a good starting place for evaluating the spread of the data.	Repeating experiments is standard scientific practice for most fields. The exceptions are usually when the scale and cost of the experiments make it impossible. For example, drug trials on a rare medical condition, large-scale sociology experiments, and astronomy observations of rare phenomena.	<ul style="list-style-type: none"> • Which wavelength of visible light emits the most heat? Make repeated measurements for each wavelength, randomize the order you conduct the wavelength experiments in, and repeat the entire set of experiments at least twice more on a different days using, if possible, different equipment. • How does caloric restriction affect the lifespan of worms? Start with a statistically large enough sample size, randomly select which worms are calorically restricted and which are allowed to eat as much as they want, repeat the entire experiment at least twice more, starting on different days, with different batches of worms.
Including covariates	Some phenomena are controlled by multiple variables. The outcome is often the sum of these covariates. When modeling these types of phenomena, or analyzing data regarding them, it is best to include all the covariates in the analysis and or model. This gives the most complete picture of what is occurring.	Including covariates is helpful if you are studying a complex phenomenon whose outcome depends on the sum of multiple factors. It is especially necessary if you can't tightly control all the variables. For example, when dealing with patients in a drug study, climate data, or analysis of food chains and ecosystems.	<ul style="list-style-type: none"> • How does drug X affect a patient's blood pressure? Start with a statistically large enough, randomly selected sample of patients. Give half of them drug X and the other half a placebo. Evaluate the results, taking into account other variables, like gender, age, and weight, known to impact blood pressure. • Evaluate how the extinction of a particular species will affect the rest of a local ecosystem. Base predictive models on historical data about the interactions between various species, starting with the one that is endangered and branching out.

Table 1. The six most common methods for maximizing the signal-to-noise ratio are listed in the table above. More details are available in the text below.

Repeated Measurements

Making repeated measurements of a single item is a powerful, but limited, technique. It is extremely helpful in cases where the measurement is challenging to make, such as in the case of observing and recording the exact instant when a liquid is completely evaporated. In these cases, averaging several measurements helps eliminate measurement error. And looking at the range and variability among the individual measurements—for example, by plotting all of them—can even help you decide if the measurement technique is adequate or simply too erratic to rely on. However, if the measurement is simple and straightforward, like weighing bags of sand on a scale to the nearest kilogram, repeated measurements add no value and instead, waste time and resources.

Increasing Sample Size

Increasing sample size is one of the most common ways to reduce experimental noise. Sample size refers to the number of items, events, or people that you make measurements or observations from in a single experiment. In general, the larger the *sample size* in an experiment, the more *power* (probability) you'll have to detect effects from changing a variable. Exactly how large should the sample size be? You can use *power-analysis* calculations to determine an effective sample size. The Science Buddies [Sample Size: How Many Survey Participants Do I Need?](#) guide includes some beginning calculations. Additional explanations and calculations can be found in the statistical textbooks like those listed in the Bibliography (look under *power analysis* to learn more). But as a rule of thumb, the smaller the expected effect, the greater the sample size you should plan on collecting. For example, if you were determining biological differences between men and women, a sample size of 10 men and 10 women would be sufficient to see that men have one X chromosome and women have two, but would be too small to conclusively determine that men are taller, on average, than women. Likewise, the more variables you are testing in a given experiment, the greater the number of samples you should evaluate

Randomization (Samples and Experiments)

Even if they recognize all the potential sources of variation, it is nearly impossible for scientists to control all factors in an experiment. Small differences in temperature, location, equipment, or other physical conditions can lead to *experimental bias* (the favoring of one outcome over another) and noise. Experimental bias and noise can be reduced by randomization. Both samples and experiments can be randomized, although it may not always be possible to use both tactics in a single science project. During sample randomization, test subjects are assigned by lottery to various control or experimental groups. For example, when studying a new diet regime, subjects would be randomly assigned to either a *negative control group*, where they are not dieting, a *positive control group*, where they use whatever diet regime is considered the gold

standard (i.e. the best diet currently known), and an *experimental group*, where they use the new diet regime. If, instead, the subjects were allowed to choose in which group they wanted to be, they might bias the results. People who willingly chose the "no diet" group might tend to eat larger meals, or the people who chose to follow the gold standard diet might be more athletic. Either of these possibilities, a tendency toward consuming more food or exercising more, might skew the results. But if the subjects are assigned randomly, such differences are likely to get distributed throughout all the experimental and control groups and thus, not noticeably skew the experimental results.

Experiment randomization can be applied in cases where there are a series of tests whose order can be determined via lottery. In these types of cases, it can be used to reduce unexpected bias in the data. For example, if the goal is to find out what level of sour flavor is tolerable for the average adult, each adult test subject would be given a series of gelatins to taste, each with a different sour intensity. The test subjects would then rate which gelatins they found tolerable and which were too sour to eat. If the test subjects were all given the gelatins to taste, in increasing order of sour intensity, the result would be an artificially inflated average sour tolerance. Why? Because systematically increasing exposure to the sour flavor temporarily desensitizes the subject's taste buds to the effects of the sourness. By randomizing the order in which each test subject tastes the various gelatins, the data is less influenced by the bias created by temporary desensitization and the resulting average is more accurate.

Repeating Experiments

Repeating an experiment also leads to an increase in the signal-to-noise ratio. Analyzing *experimental repeats* diminishes the chance that spurious effects (like a slightly raised ambient temperature or a machine whose readings are too high) are driving the conclusions. Data from samples are collected together in a single experiment; a repeat of an experiment needs to be *independent*, meaning as many of the experimental parameters as practically possible should be changed: different samples, different machine, different day, different experimenter etc. Three repeats of an experiment is generally considered the minimum. Why? There are two reasons, the first has to do with the fact that three repeats ensures a two-thirds (66%) probability that the averaged results are more accurate than a single experiment. Two-thirds may not seem like a lot, but repeats have a diminishing return—more than three and you have to do a lot more repeats to make a major increase in confidence. Even with 500 repeats, there is still a small chance that a single trial will just happen to be closer to the true value than the average. See Table 2, below, for details. The second reason is that with three repeats, you have a good basis for graphing and using statistical descriptions, like mean and standard error of the mean, to evaluate your data and see if the results are robust enough to make a conclusion from, or if you need to gather more data. In some cases, repeating an experiment is not possible due to resource constraints. For example, a biological survey of a large track of land, like the Amazon rain forest, would only be carried out once. When repeats are not going to be possible, it is critical to be sure the sample size is sufficiently large.

# of Experimental Repeats	% Chance the Average of the Repeats is More Accurate Than a Single Trial
2	60.8
3	66.7
4	70.5
5	73.2
10	80.5
20	86.0
40	90.0
100	93.7
162	95.0
500	97.2

Table 2. Repeating an experiment a few times results in a large increase in the statistical chance that the average of the repeats is more accurate than a single trial of the experiment, but subsequent repeats have diminishing returns. (Table adapted from Gauch, 2006. See original text for underlying theory.)

Including Covariates

Many natural systems and scientific phenomena are the sum effect of many factors. These factors, called *covariates* because they "vary together," collectively control the final outcome. Although scientists are often interested in assessing how changing a single one of the factors will affect the whole system, it can be impractical, or even impossible, to set up an experiment where just one variable can be changed and evaluated. For example, if you wanted to predict how building a new car-manufacturing plant would affect the local air quality, one way would be to just determine how much air pollution the factory would contribute. But this model is imprecise. There are other related events that might occur when a new

factory is built. For example, the factory would create jobs, and more people might move to the area to take advantage of those jobs. These people would buy local homes, drive cars, start related industries, and so forth. All these events would also impact local air quality. So, a more-accurate evaluation would take into account as many of the covariates as possible.

Taking covariates into account can also help increase your power of detecting a change. For example, say you were conducting a study on the ability of a new drug to lower cholesterol. Cholesterol levels are determined by a large number of factors, including: gender, age, family history, diet, physical activity, and weight. In a study with mice, you could control for all of these factors—you could have mice with identical genetics, all of the same age and gender, that are fed the same diet, that weigh the same amount, and perform the same exercise regime. But it would be impossible to do a similar fully controlled study with humans. And each factor you try to control, the fewer people would be available to your study and the more difficult it would be to recruit subjects. An alternative is to limit only some of the variables, and measure the remaining covariates in order to factor them into your final data-analysis model. Using the model, you can mathematically subtract out the effects of the covariates and still see the effects of the variable in which you're interested: the cholesterol-lowering drug.

Bibliography

These resources provide additional information about how to design experiments and increase the signal-to-noise ratio in scientific data:

- Anderson, M.J. and Anderson, H.P. (1993, July/August). Applying DOE to Microwave Popcorn. *PI Quality*. Retrieved August 25, 2009, from <http://www.statease.com/pubs/popcorn.pdf>
- Gauch, H.G., Jr. Winning the Accuracy Game. *American Scientist* 94, 2, 133.
- Khare, R. (n.d.). *Three Romeos and a Juliet: Our Early Brush with Design of Experiments*. Retrieved August 25, 2009, from <http://www.symphonytech.com/articles/romeos.htm>
- Trochim, W.M.K. (2006). Introduction to Design. *Research Methods Knowledge Base*. Retrieved August 25, 2009, from <http://www.socialresearchmethods.net/kb/desintro.php>
- Wilson, E. B. *An Introduction to Scientific Research* New York: Dover Publications, Inc.



It's *free!* As a member you will be the first to receive our new and innovative project ideas, news about upcoming science competitions, science fair tips, and information on other science related initiatives.

[Science Fair Project Home](#) [Our Sponsors](#) [Partners](#) [About Us](#) [Volunteer](#) [Donate](#) [Contact Us](#) [Research Grants & Outreach](#) [Site Map](#)

[Science Fair Project Ideas](#) [Science Fair Project Guide](#) [Ask an Expert](#) [Blog](#) [Teacher Resources](#) [Parent Resources](#) [Student Resources](#) [Science Careers](#) [Join Science Buddies](#)

[Privacy Policy](#)

Copyright © 2002-2010 Science Buddies. All rights reserved.
Reproduction of material from this website without written permission is strictly prohibited.
Use of this site constitutes acceptance of our [Terms and Conditions of Fair Use](#).