I. Introduction to Imperfection

If the same thing is measured several times, in an ideal world the same result would be obtained each time. In practice, there are differences. Each result is thrown off by *chance error*, and the error changes from measurement to measurement. *Freedman et al, p 91,*

II. What is Chance Error?

Multiple measurements of virtually anything, for example basketball players heights, the amount of money people win playing dice in Las Vegas, your scores on a test, often follow a normal curve. The average of these measurements estimate its true value (e.g. height, money, score) and the SD of these measurements estimate the chance error of the measurements with respect to the true value. Bias arises if all the measurements are consistently different from the true value in some way.

Example: *Measurements by NIST*1 Eleven measurements of the mass of the *same* object, taken with a much care as possible by the National Institute of Standards and Technology (NIST), gave the following results (in grams):

9.9995992, 9.9995947, 9.9995978, 9.9995925, 9.9996006,
9.9996014, 9.9995985, 9.9996008, 9.9996027, 9.9995929, 9.9995988.

(a) **True** / **False** One particular measurement is 9.9995985.

(b) **True** / **False** The measurements vary even though great care was taken to measure the object with accuracy.

The difference between the true mass and a measurement of this mass is called the *chance error*.

The example of weights makes the point that even the most careful measurements are often different.

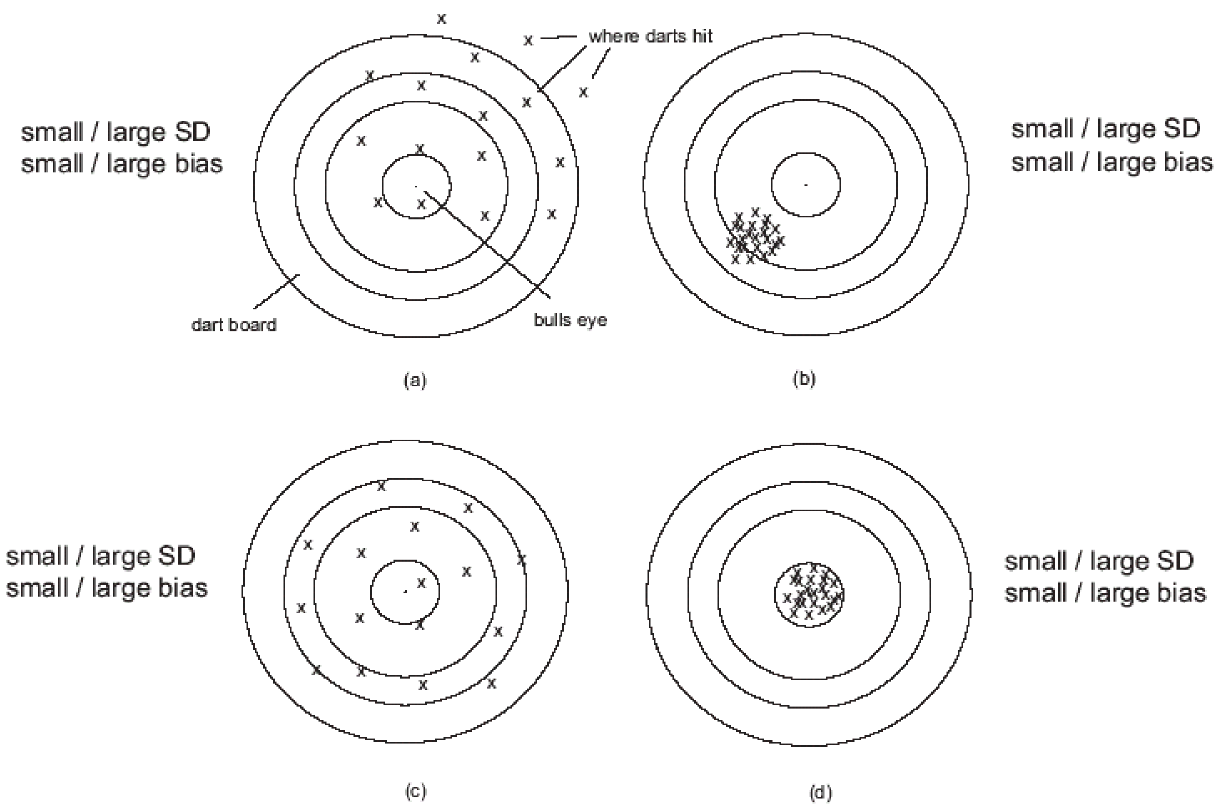
Realize that any measurement follows the formula:

observed measurement = true value + bias + chance error.

By taking a series of measurements, we can compute the SD for the series, and, in turn, estimate the likely size of the chance error. Sometimes there are observations, called *outliers,* that are several SDs away from the average. Sometimes a reason can be found for outliers, and you can justify excluding them from the data set. For example, perhaps the worker recalls that a certain observation was made with a dirty scale pan, which would add some weight to the measurement. You could reasonably exclude the measurement on that basis. If, however, you cannot find a reason to exclude the outlier, you should include it. The presence of many outliers is an indication that the data may not follow the normal distribution.

Imagine that all the measurements were made with a dirty pan. In that case, the average would be systematically higher than the true value. Such systematic error is called *bias.* Detecting bias is difficult: you must know the true value to know that the measurements are biased. Ideally of course, we want bias to be equal to zero and the chance error to be equal to zero. Change error will always exist, but we can do things to eliminate bias.

Chance Error and Bias: Darts Where our darts land on the board will represent the different values of our measurement. The goal in our game of darts is to try and hit the bulls eye. When our darts land on or very near the bulls–eye, this is equivalent to our measurement being equal to or near equal to the actual unknown true measurement.



How does this apply to us? Consider the idea that the bulls-eye represents the true parameter and the "darts" represent outcomes from samples. For example, suppose is known that 48% of the registered voters in Los Angeles, California are registered as Democrats (this is a parameter). To test a new telephone sampling method, we call 500 Los Angeles County voters and ask their party. We do this 5 times. The results are 49.2%, 48.9%, 48.5%, 48.4% and 47.8% Democratic. The sampling method appears to have:

(i) high bias and high chance error
(ii) high bias and low chance error
(iii) low bias and high chance error
(iv) low bias and low chance error

In practice then, multiple measures (from samples) are used to estimate all kinds of parameters such as election turnout predictions, unemployment rate, air quality, water quality, percentage of new cars with problems.

III. Bias

Bias is **systematic error** (i.e., something that, on average, cause the true value to either be under- or over-estimated). Bias cannot be removed with repeated measurements. Having more measurements will not eliminate bias.

Example: Suppose it is advertised that the toner cartridge made by a company lasts 5.5 months under regular use. To test the advertising claim, nine toner cartridges are sampled at random. It is found that the nine toner cartridges lasted the following number of months:

4.9, 4.7, 4.8, 4.8, 5.0, 4.7, 4.9, 4.8, 4.8

The sampling method appears to have (circle one)
(i) high bias and high chance error
(ii) high bias and low chance error
(iii) low bias and high chance error
(iv) low bias and low chance error

Bias can be avoided by doing everything possible to make sure that the measurement is fair.

IV. Outliers

Outliers are extreme values and they have two possible sources:

1. a variable may simply have lots of variability
2. some type of problem (e.g., typo, misread, mismeasurement etc.).

A. How are outliers defined?

1. No definite answer, it's an art.
2. Many researchers simply adopt some type of criterion (e.g., if a value is > Mean + 3 Z scores or a value is < Mean – 3 Z scores, then the value is considered to be an outlier) *before* they start analyzing their data.
3. Descriptive statistics (especially histograms) are critical to get some sense of what a particular set of data look like.

B. How are outliers handled?

1. Again, no definite answer.
2. Some people replace outliers with means (or medians).
3. Some people simply remove them. (see "outlier history" below)

The effect of outliers can be minimized by taking many measurements (i.e., using a large sample) and then using the mean of these measurements (i.e., the sample mean) to estimate the population mean.

V. Outlier History – The Ozone Layer & CFCs

Satellite measurements of ozone started in the early 70's, but the first comprehensive worldwide measurements started in 1978 with the Nimbus-7 satellite. Chloroflourocarbons were created in 1928 as non-toxic, non-flamable refrigerants, and were first produced commercially in the 1930's by DuPont.

Worldwide consumption in 1988 was estimated at over billion kilograms. In 1974 M.J.Molina and F.S.Rowland published a laboratory study demonstrating the ability of CFC's to catalytically breakdown Ozone in the presence of high frequency UV light. Further studies estimated that the ozone layer would be depleted by CFC's by about 7% within 60yrs and based on such studies the US banned CFC's in aerosol sprays in 1978. Slowly various nations agreed to ban CFC's in aerosols but industry fought the banning of valuable CFC's in other applications. A large shock was needed to motivate the world to get serious about phasing out CFC's and that shock came in a 1985 field study by Farman, Gardinar and Shanklin. Published in *Nature*, May 1985, the study summarized data that had been collected by the British Antartic Survey showing that ozone levels dropped to 10% below normal January levels for Antarctica. The authors had been somewhat hesitant about publishing because Nimbus-7 satellite data had shown no such drop during the Antarctic spring. But NASA soon discovered that the spring-time "ozone hole" had been covered up by a computer-program designed to discard sudden, large drops in ozone concentrations as "errors". The Nimbus-7 data was re-analyzed without the computer program and evidence of the Ozone-hole was seen as far back as 1976. Numerous studies since then have confirmed both the Antarctic hole, as well as an overall global decrease in Ozone.

Moral: Don't just toss out outliers. They may be the most valuable members of a dataset!