Chauvenet's Criterion

Everybody makes mistakes. Of course, when you are making observations and measurements you want to minimize the number of mistakes but they regularly occur. For example, you could have transposed a couple digits in your notes, maybe you hit the wrong key on the keyboard, maybe the instrument went nuts after you nudged it, or maybe it is somebody else's fault. In the course of inquiry, everybody comes up with some data they think should be excluded from their study.

If you have a measurement that is radically different than the others in a series of measurements, you have to decide why it is different. That is, you have to decide if the anomalous measurement is the result of some mistake or whether it is representative of the population being sampled or measured and should be included with all the other measurements. This is one of those situations where good, accurate, and legible field and/or lab notes will prove their usefulness – look at those notes and see if you recorded anything peculiar about the suspect measurement. If the anomalous result is from a measurement which can easily be reproduced, like a measurement of gravity at a known location, go back and double check it.

As Taylor[1] points out, omitting any data is controversial, and some scientists think you should never do so. However there is a simple, established test, Chauvenet's criterion, you can consider employing when you have an anomalous measurement. Taylor's statement of Chauvenet's criterion is:

> "If you make N measurements x_1, x_2, …, x_n of a single quantity x, and if one of the measurements (x_{sus}, say) is suspiciously different from all the others, Chauvenet's criterion gives a simple test for deciding whether to reject this suspect value. First, compute the mean and standard deviation of all N measurements and then find the number of standard deviations by which x_{sus} differs from x_{mean},
>
> $$t_{sus} = \frac{\left| x_{sus} - x_{mean} \right|}{s_x}$$
>
> Next, find the probability (assuming the measurements are normally distributed about x_{mean} with width of s_x) of getting a result as deviant as x_{sus}, and hence, the number of measurements expected to deviate this much,
>
> $$n = N \text{ x } \boldsymbol{Prob}(\text{outside } t_{sus}\mathbf{s})$$
>
> If $n < 0.5$, then according to Chauvenet's criterion, you can reject the value x_{sus}."

Suppose you have an anomalous value two standard deviations from the mean. Ninety five percent of your Gaussian distributed samples should be within two standard deviations of the mean. The probability of being outside two standard deviations is one minus the probability of being with two standard deviations, or 5%:

---

[1] Taylor, John, R., 1997, An Introduction to Error Analysis: The Statistical Study of Uncertainties in Physical Measurements, University Science Books, 327 p. Recommended!

Prob(outside 2 standard deviations) = 1 – ***Prob***(within 2 standard deviations).
Prob(outside 2 standard deviations) = 1 – 0.95
Prob(outside 2 standard deviations) = 0.05

That is, you expect one in twenty measurements to exceed two standard deviations. If you make twenty measurements and one or two exceed the second standard deviation, you cannot reject them.

If you only made five observations, then your expected number of measurements outside of two standard deviations is:

Expected = (number of measurements) x ***Prob***(outside two standard deviations)
Expected = (5) x ***Prob***(0.05) = 0.25

So with your five measurements you only expect one quarter of a measurement to exceed two standard deviations. Thus you might choose to apply Chauvenet's criterion and exclude one anomalous observation from a set of five; it is up to you to make, explain, and justify such decisions. Read Taylor (1997) for details on this an a large number of other interesting topics concerning errors.