MICROPIPETTE CALIBRATIONS

I. ABSTRACT

The micropipette set is a basic tool in a molecular biology-related lab. It is very important to ensure that the micropipettes are properly calibrated, so that you have confidence in the volumes dispensed.

II. OBJECTIVES

- Determine whether the micropipettes in your set dispense the desired volumes
- Determine the spread of values (i.e., error/uncertainty) in dispensed volumes
- Review statistics of fitting to noisy data and apply to measurements here

III. MATERIALS AND EQUIPMENT

- one set of micropipettes (P-1000, P-200 and P-20 or P-10) and micropipette tips
- 50 ml falcon tube containing distilled water
- analytical balance with sub-mg resolution
- weigh boats

IV. PROTOCOLS

1. Set the first micropipette to dispense the maximum volume suggested.
2. Write down this volume (with the appropriate units) in your notebook.
3. Place an empty weigh boat in the analytical balance, close the doors, and tare the balance (adjust zero reading for this weigh boat).
4. Using the micropipette, transfer the set volume of water from the tube of distilled water to the weigh boat in the balance. (See Fig. 1 for how to use a micropipette.)
5. After closing the doors on the balance, read and note the mass of water (with the appropriate units) you have dispensed.
6. Repeat steps 3–5, for a total of 10 readings for this pipette volume.
7. Make single measurements for the rest. Measure 2–3 volumes per pipette. Then measure the other micropipettes in your set.

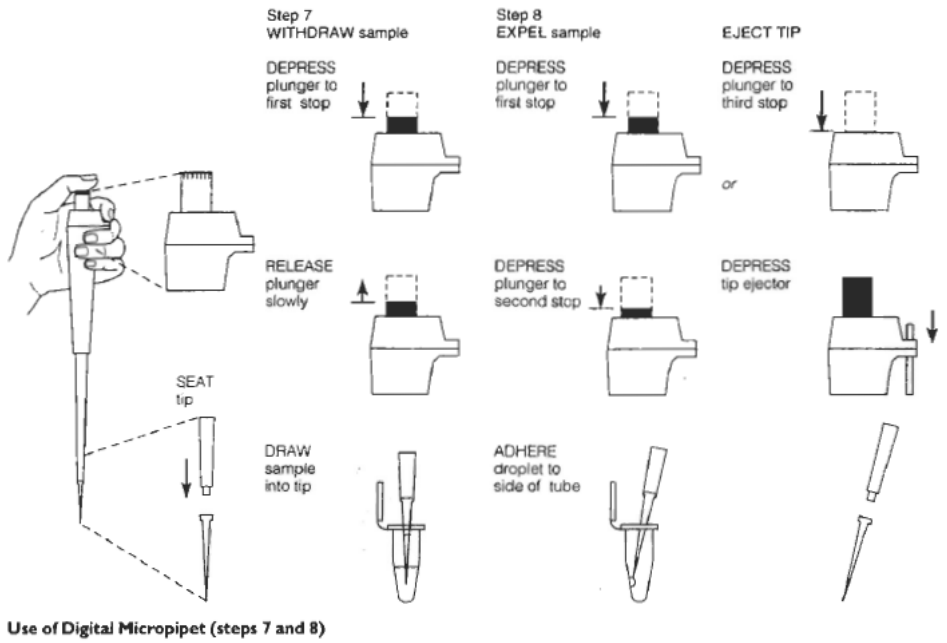

FIG. 1. Figure showing proper use of micropipette. (From *Laboratory DNA Science*, by Bloom, Freyer and Micklos, The Benjamin/Cummings Publishing Company, Inc., 1996, p. 8.)

V. ANALYSIS

The first issue is to understand the statistical errors that occur when you make a pipette measurement. Recall that *statistical errors* are different for each measurement. They often reflect the combined impact of many different factors that are each hard to control. *Systematic errors*, on the other hand, give the same (incorrect) result for each measurement. A classic source is miscalibration.

To determine the properties of statistical errors, you need to do repeated measurements. If possible and practical, automation is the way to go. Here, with manual pipettes and limited time, we adopt the old-fashioned, manual method where we manually repeat a small number (e.g., 10) measurements and look at their statistical properties. The questions we want to ask include

- Are the errors at least approximately Gaussian? One test is to make a histogram and see whether the result is at least roughly a bell-shaped curve. With only 10 measurements, this will be very rough.

- Given at least an approximate Gaussian distribution, we characterize it by its standard deviation (or variance).

- Finally, a very common situation is to have most of the measurements fall in a narrow (Gaussian-distributed) range but to have an occasional outlier ("blunder") that is so large that it cannot be explained as a rare instance of the normal Gaussian errors.

Our goal then is to figure out whether (1) errors are roughly Gaussian; (2) their standard deviation is independent of the volume measured, proportional, or something more complicated; and (3) the proportion of outliers. To make this feasible with manual measurements, each group should measure a different volume, and all groups should pool their data together to address the above questions. To get you started, Fig. 2 shows a typical measurement of 10 errors from a Gaussian distribution. The errors were drawn from a Gaussian distribution of variance (and, thus, standard deviation σ) =1. The estimated $\sigma = 0.8$. Note that you have to play around with binning to get an appropriate curve. Too narrow and the results will be very noisy; too wide and you lose features of the distribution. With only 10 observations, you may not get as pretty a distribution as this one.

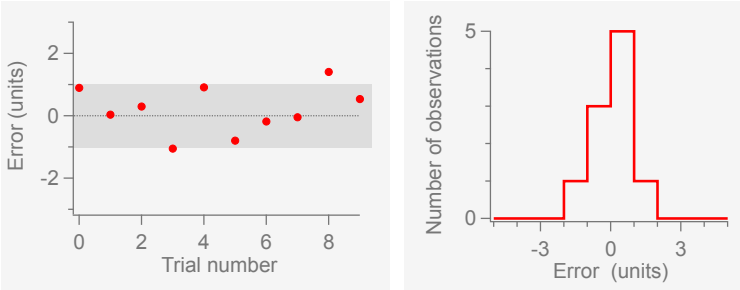

FIG. 2. Determining error distribution. Left: statistical errors from 10 measurements. Right: Histogram of these measurements.

For a single group, for each of the 10 measurements, determine

1. The mean mass of water dispensed.
2. The standard deviation of a *single* mass measurement.
3. Assess, at least qualitatively, whether the 10 measurements are distributed as a Gaussian. An easy way is to plot the points and the mean and to give the points error bars. Recall that about $^2/_3$ of the points should lie within one standard deviation and that the others should lie outside (but probably not too far). This test can be done more formally, e.g., using the χ^2 statistic, but we will save that for the actual curve fits we do, below. Least squares fits are not incredibly sensitive to the exact noise distribution, so as long as the errors are roughly Gaussian, least-squares is a reasonable way to analyze the data. In this part, you are mainly looking for obvious outliers or "bad points."

When you combine your results with those of the other groups, you can decide on an error model (e.g., Gaussian with $\sigma \propto$ volume (with an estimate of the proportionality constant).

We will assume that this standard deviation holds for the other measurements. Why is this reasonable? Then, for each micropipettor,

4. Plot the mass of water dispensed versus set volume on the micropipette. Include error bars whose length is given by $\pm$ the standard deviation. We expect (or at least hope) that this relationship will look linear.

5. Using your favourite curve fit program, find the slope and its standard deviation. Decide whether to fit the intercept or force it to pass through zero.

6. Assess the fit quality three ways:

   (a) By eye: does the line fit the points as well as you expect?

   (b) By the normalized residuals: do they look like they are drawn from $\mathcal{N}(0,1)$? (See below.)

   (c) From the χ^2 of the fit. Be sure to give the curve-fit program the standard deviation of each measurement value so that it can make a normalized evaluation of the error.

VI. ITEMS TO INCLUDE IN LAB WRITE-UP

In your lab report, please do the following:

- show raw data, e.g., in a computer-table form; You can include the data in your lab report as a figure generated from the data analysis program (so that you don't have to retype them).
- show the results of the analysis;
- discuss possible sources of measurement noise ("experimental error");
- discuss the reliability of your micropipettes.

VII. STATISTICS REVIEW

Reference: D. S. Sivia, *Data Analysis: A Bayesian Tutorial*, Oxford Univ. Press, 2nd ed., 2006. Read Chapters 1–3.

For this particular lab, you should know how to calculate the mean and standard deviation of a set of measurements and how to fit noisy data to a straight line via the least-squares technique. I hope that you have seen those techniques in previous courses. But reading Sivia will tell you WHY these are the right things to do.

The end result of your analysis of pipette data should look something like Fig. 3. One slightly non-standard request is to plot the *normalized* residuals, rather than the standard residuals. This is good practice. To understand why, we assume that we have N data pairs (x_i, y_i), for $i = 1, \ldots, N$. The x are assumed noiseless, for simplicity (*independent* variables). The y are the noisy measured data, and we assume that they are related to the x by

$$y_i = f(x_i) + \xi_i \,. \tag{1}$$

In this lab, we expect $f(x) = a_0 + a_1 x$ is a straight line with unknown intercept a_0 and unknown slope a_1. The term ξ_i represents the noise, which we assume to be distributed as a Gaussian distribution of mean 0 and standard deviation σ_i. The subscript reminds us that the noise can differ for each measurement. This will happen, for example, if you use a different number of trials; we do not in this lab. In fancier mathematical notation, we write $\xi_i \sim \mathcal{N}(0, \sigma_i^2)$. The tilde, $\sim$, means "is distributed as", and the notation $\mathcal{N}(\mu, \sigma^2)$ refers to the *normal* distribution of mean μ and variance σ^2. ("Normal" is another term for "Gaussian.") It is conventional to make the last argument the *variance* (square of the standard deviation), which explains why we write σ_i^2 and not σ_i.

The normalized residuals are then defined as

$$\varepsilon_i = \frac{y_i - f(x_i)}{\sigma_i} \,. \tag{2}$$

It is nice to write them in this way because $\varepsilon_i \sim \mathcal{N}(0, 1)$. That is, the normalized residuals should be Gaussian distributed with mean 0 and standard deviation 1. Often, you can tell whether this is the case just by looking at the residual plot. Of course, the χ^2 statistic is a more "professional" way of assessing whether a fit is good statistically. (But if the fit fails the qualitative test, there is not much point in calculating the significance of χ^2. Rather, it's a more stringent test that is useful IF the data look more or less reasonable by eye.) Note that many analysis programs, such as Igor, supply the unnormalized residuals, $y_i - f(x_i)$. You can easily normalize them by dividing by the individual standard deviation.

In your analysis, both by looking at the residuals and by the more formal χ^2 test, you will assess whether your fit is good and, hence, whether the linear model of pipettor measurements is reasonable.

A couple of final observations: The procedure suggested here is to first assess whether the statistical errors of a measurement are distributed as a Gaussian. We use a couple of qualitative tests: the proportional of points that fall within $\pm 1\sigma$ and the shape of a histogram. More "professional" tests are available to make a sharper assessment. For example, one can fit the histogram to a Gaussian. Alternatively, one can calculate statistics such *skewness* (zero for a distribution such as a Gaussian that is symmetric about the mean) and *kurtosis* (related to the fourth moment and describing whether the distribution is "sharper" or "flatter" than a Gaussian) and compare to what is expected for a Gaussian.

Another important observation is that the average of several measurements has a distribution that approaches a Gaussian (as the number of measurements N becomes large), even if the distribution of measurements is far from a Gaussian. The *Central Limit Theorem* gives conditions for this to happen, and it is a powerful tool for making sure that errors are Gaussian. The problem is that you need to repeat measurements for every data point. If measurements are automated, this is not difficult and is the way to go. Our procedure here is a compromise for a case where taking many measurements for each point would be very slow.

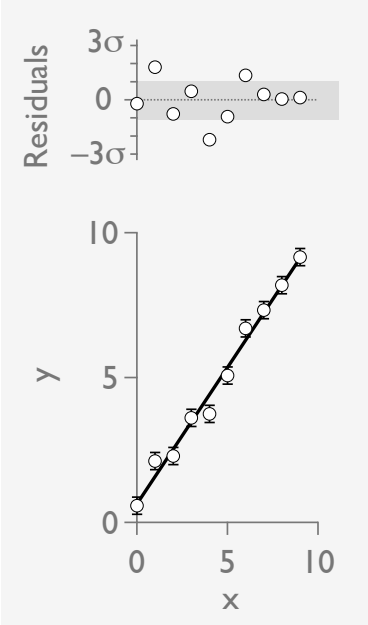

FIG. 3. Simulated data from a linear relationship, showing normalized residuals. Notice that about $^2/_3$ of the data points lie within 1σ of the best-fit line (and, thus, about $^1/_3$ do not).

