

Estimation of Gravitational Acceleration g Using a Simple Pendulum: A Regression Approach

William C. Evans August 2023

INTRODUCTION

A method is presented for estimating the gravitational acceleration g using a simple pendulum. This method uses several pendulum lengths and a regression process that could in principle be done on a calculator, but with some extra effort (it requires calculations that are not provided in the device, and must be separately programmed). Thus, this method is intended for implementation on a computer, and that code is illustrated here with MATLAB, a free, open-source version of which is available, called "Octave."

PROCEDURE

Set up the pendulum apparatus, having at least 5 string lengths (L) available. Lengths might be 0.1, 0.2, 0.3, 0.4, 0.5 meters, but these need not be the exact lengths. Spreading the lengths out more or less evenly across a 0.5m maximum is desirable. Carefully measure the pendulum length from the suspension point at the top of the string *down to the center of mass* of the pendulum "bob" (a fishing-line sinker works well). Failing to include that extra length will introduce a systematic error; this has been observed to happen, and it will bias the results.

Displace the pendulum bob, holding the string taut (straight) at an angle of no more than about 10-20 degrees from the vertical. Try to *use the same angle each time*. Measure this initial displacement angle θ_0 once, with a protractor; we will need to use this value in subsequent calculations. This angle is important to estimate, since it affects the pendulum's period T , but small errors in θ_0 will not be significant. Release the "bob" without imparting a velocity; let go of it, don't throw it.

Observe and record neatly in a table the time needed for the pendulum to make n complete cycles, at each of the several pendulum lengths. Five cycles is a good choice; if n is too large, damping could affect the data. Repeat this measurement, at each pendulum length, at least $N=5$ times. (n is the number of cycles per measurement; N is the number of replicated measurements.) Do not divide the periods T by the number of cycles n at this point; this will be accounted for later.

DATA ANALYSIS

We will use a special form of linear regression to estimate g from this data. To see how this arises, consider the usual relation for the pendulum period T , modified for the case of n cycles observed, and accounting for the effect of θ_0 (the usual T equation assumes a "small" initial displacement angle, usually less than ~ 5 degrees; see, e.g., Sears and Zemansky 10th Ed. p408 for discussion of this displacement-angle correction factor):

$$T = 2 \pi * n * \sqrt{L/g} * (1 + \frac{1}{4} \text{Sin}^2(\theta_0/2)) \quad (1)$$

which can be written as

$$T = [2 \pi * n * \text{AngleCorrectionTerm} / \sqrt{g}] * \sqrt{L}$$

or

$$T = b_1 * \sqrt{L}, \quad \text{where } b_1 = 2 \pi * n * \text{AngleCorrectionTerm} / \sqrt{g} \quad (2)$$

This is the equation of a linear function, in fact a simple straight line, with zero intercept and a slope b_1 . Here the y -variable is the **mean** of the N observed n -cycle periods, at each of the several lengths L , while the x -variable is the square root of the length L . Note that we are transforming the *independent* variable, L , rather than squaring the *observed* variable, T . **It is good statistical practice to avoid nonlinear transformations of an observed random variable; such transformations introduce unnecessary complexities into the data analysis.** Linear operations such as multiplication or division by a constant do not cause any difficulties, however.

In regression analysis it is shown, and simulation studies verify, that a "better" (more precise) estimate of the slope b_1 will be obtained if there is no intercept in the "model" Eq(2) *when it is not needed*. We will call this form of linear regression a "zero-intercept regression" or ZIR. To find an estimate of g ("g-hat" means estimate), once we have estimated b_1 from the data using a ZIR,

$$\hat{g} = (2 \pi * n * \text{AngleCorrectionTerm})^2 / b_1^2 \quad (3)$$

This provides a "point estimate" for the gravitational acceleration g . Next we need this estimate's uncertainty, which is found using the usual "propagation of error" (POE) calculation, given as Eq(2) in the "Experimental Data Analysis" PDF in this series (see the Wikipedia article *Experimental Uncertainty Analysis* for a derivation of that equation).

One factor in this uncertainty analysis is the initial angle θ_0 . We will estimate it at the beginning of the experiment, and that estimate may have a *systematic* error, since it is difficult to position and read a protractor near the pendulum string. Then, during the experiment, we need to repeatedly displace the pendulum by this same angle, as best we can, for each replicate, at each length. This cannot be done exactly, so that there will be some variation in the initial angle (*random* error). That in turn will contribute a small amount, through the *AngleCorrectionTerm*, to the variability in the observed periods T , along with the usual timing "errors." However, all the T variability, regardless of its source, will be manifested in the ZIR b_1 uncertainty. Thus, for our POE calculation we don't need to explicitly include the variability in the initial displacement angle; this is implicit in T . Then b_1 will be the only random variable in Eq(3), and using the POE equation we will find:

$$\sigma(g\text{-hat}) = 8 \pi^2 * \text{AngleCorrectionTerm}^2 * n^2 / b_1^3 * \sigma(b_1) \quad (4)$$

where $\sigma(b_1)$ is provided by the ZIR. This estimator has been checked with numerical simulations, and it does agree with the observed variability of the g estimates. The observed standard deviation of one run of 10000 of these g estimates was 0.2042, while the mean value of the corresponding 10000 POE standard deviations, using Eq(4), was 0.2045. These g estimates also appear to closely follow a Normal distribution (Fig. 1), which means that constructing confidence intervals using a t-statistic will be appropriate (A confidence interval is a range of values within which it is expected that the true value of a quantity will be found; see the "Confidence Interval Illustration" PDF in this series for more on this statistical concept, including why the "t-statistic" is to be used.)

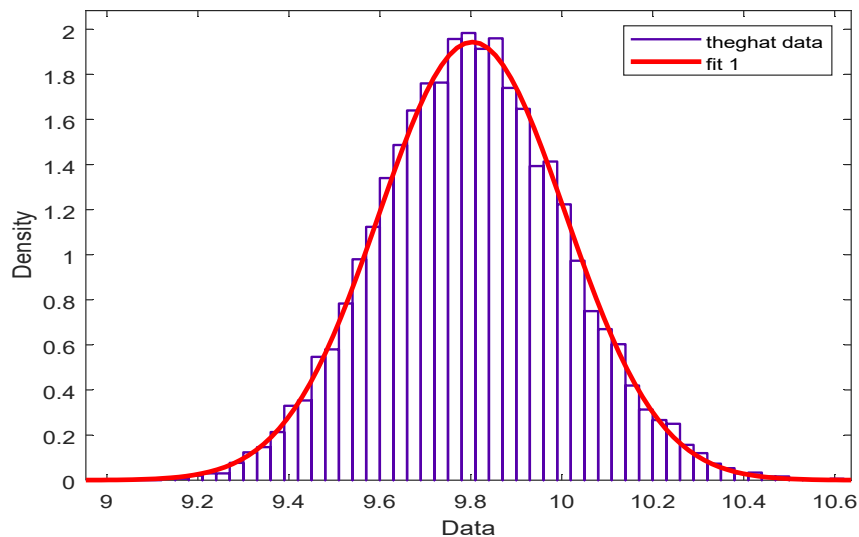


Figure 1. Histogram of 10000 ghat estimates, found using ZIR. Normal PDF shown (red line).

EXAMPLE DATA

Below is a synthetic dataset for a pendulum experiment, with the time to complete $n = 5$ cycles measured. The initial angle was 45 degrees; this is larger than recommended above, but is used to illustrate that the *AngleCorrectionTerm* in Eq(1) does correct for this relatively large initial displacement angle. The average T values at each L are shown as well.

L, m	T, sec					Mean T
0.10	3.39	3.21	3.62	3.48	3.63	3.466
0.20	4.59	4.89	4.41	4.65	4.92	4.692
0.30	6.25	5.40	4.87	5.79	5.50	5.564
0.40	6.69	6.52	6.68	6.75	6.51	6.632
0.50	7.13	7.75	7.17	7.57	7.20	7.364

COMPUTER ANALYSIS

The data analysis can be done using a TI-84 series calculator, but this is tedious and also requires programming certain statistical calculations that are not available (the TI "code" for those calculations is provided in the "Experimental Data Analysis" PDF). So, if possible, it is preferable to do the analysis with a computer, using code written in R, or MATLAB/Octave. Octave is a free, open-source version of MATLAB. R is also free, and widely used for statistical data analysis. For the synthetic dataset above the results from MATLAB are

Estimated-g: 9.750 +/- 0.175 Estimated-g 95% confidence bounds: 9.265 10.235 t-stat 2.776

which has a remarkably small **relative error** (uncertainty times t-statistic, divided by the point estimate) of about 5%. This is a consequence of the size, which was apparently a bit unrealistically small for student measurements, of the "noise" added to the T values in generating the synthetic data. Note that the "confidence interval" bounds for the g estimate include the known value of 9.8 m/s^2 .

One advantage of computer use vs. calculators is the availability of far superior graphics (if one takes the time to code this, of course). These plots can be helpful in detecting systematic errors in the data collection. Figures 2 and 3 are examples of the graphical output from the MATLAB program that processed the example dataset above (Fig. 2), and a different dataset that had a systematic error (in θ_0) deliberately introduced (Fig. 3).

Below is the MATLAB code for the ZIR estimation. The method of input (e.g., to read a text file of the data, or perhaps just to enter it interactively), the graphical and/or tabulated output, etc. can be developed as required, but this estimation process may not be readily available, so it is provided here:

```
xdesign = [ sqrt(L') ];  
[beta, betasigma] = lscov( xdesign, Tmean' );  
theslope = beta(1);  
slopesigma = betasigma(1);  
yresid = Tmean' - xdesign*beta;
```

L is a vector of the pendulum lengths, $Tmean$ is a vector of the average T at each length. $beta$ is a vector (here of size one) of estimated regression coefficients; $betasigma$ are their uncertainties. This code fragment has been tested in Octave; it works. It may be possible to install Octave on the school computer system, for the students to access.

CONCLUSION

The estimation of the gravitational acceleration g with a simple pendulum using several lengths has been illustrated. Design of an actual lab exercise to be used in class can be developed using this example as a template. It is suggested that the students process the synthetic data given here a day or two before the actual lab, as homework. Then on lab day they will have seen what their data should look like, and will have practiced the data processing.

A key point here is the *emphasis on the uncertainty in the estimate*; this is often not considered. Generally, science lab grading is done based only on a "percent difference" between the student's estimate and the known value. But sloppy, inattentive work will lead to large variability in the data, while the point estimate may still, accidentally, turn out to be close to 9.8 m/s^2 . That lucky result leads to a small percent difference, and a good grade. This is not fair to other students who worked carefully and thus had far less variability in their data, even if their point estimate happened to be a bit farther off. The uncertainty in the latter group's estimate of g will be considerably smaller than that of the former group, and this more attentive work should be rewarded. **Grading should be based on BOTH accuracy (percent difference) and precision (relative error)**. The grading scales for these two measures might be based on the instructor's results when doing the lab, indicating what "good quality" results should look like, or perhaps a sliding scale ("curve") using all the student data/results.

FIGURES 2, 3: NEXT PAGE

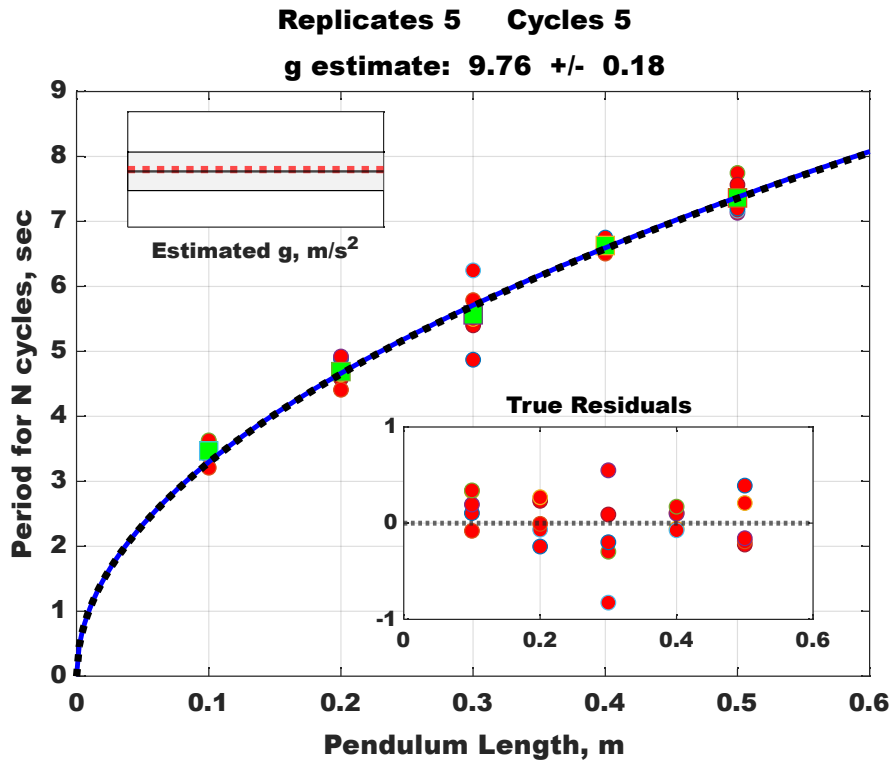


Figure 2. Results for example dataset on page 2 above, MATLAB. Dashed curve is for $g = 9.8 \text{ m/s}^2$ exactly; the blue curve is for the ZIR-estimated g . The horizontal scale here is linear; if we had plotted the square root of the length, the graph would be a straight line. "True" residuals are for the correct- g curve rather than the estimated- g curve. The box at top left illustrates the 95% confidence band on the g estimate; the dashed line is exactly 9.8 m/s^2 . These results are excellent, showing that the initial-angle correction works well.

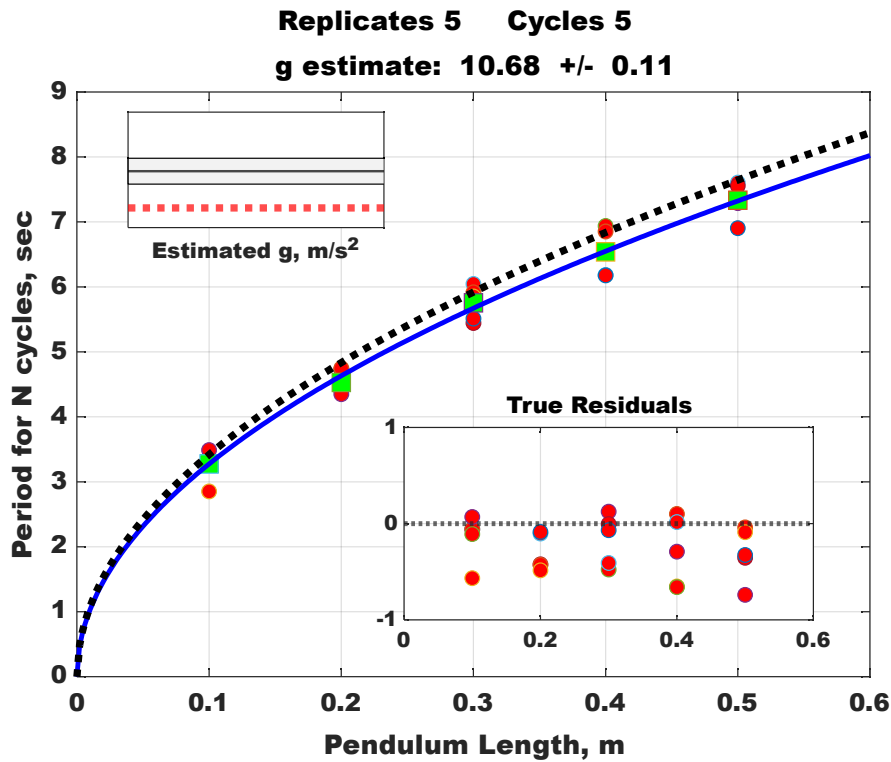


Figure 3. Example of graphical output for biased data; a deliberate systematic error in measuring the initial displacement angle. Note that the confidence region for the estimated g does not include the correct value (dashed line), the difference from the known curve is conspicuous, and the residuals are trending below the zero level. Interestingly, it took a relatively large error in the initial angle to make these results noticeably different from those in Fig. 2.