Mike Sutherland (55505225)
Instructor: Dr. Marco Maggia
TA: Daniela Ruiz

Lab Report 1 – Accuracy, Precision, Significant Figures

Summary (Answer to Question 1)

The experiment was the indirect measure of a water flow rate by timing the flow of a volume of water into a container. Four different combinations were tested: high flow rate and small volume, high flow rate and large volume, low flow rate and large volume, and low flow rate and small volume. Additionally, another set of experiments measuring the same quantity were performed with a different apparatus.

The purpose of the lab was to shed light on the impact of experiment design on precision and accuracy of the ensuing data, and to understand the full meaning of precision and accuracy in a lab context. The lab explored how the presence of outliers can alter the perceived variance of a measure and tests for when to remove such outliers.

Questions

Question 1: Write a summary of the experiment, results, and explain how this lab benefits analysis and interpretation of experimental data in general.

(Answered in Summary Section)

Question 2: Plot the frequency distributions of Q. Are there any outliers? If so, re-plot the frequency distributions without outliers. Why are the distributions different?

We can see the graphs in the Graphs section of the report. We notice that the distributions are different. For the fast pour, there are no outliers, so the two charts (at left on both ensembles) are the same, because no outliers are discarded from the distribution. However, on the right, we see that there are outliers, specifically on the large, slow pour. The distribution for that pour is “tighter” (i.e. the histogram is narrower), because the presence of the outliers increases the variance for the sample. This makes sense; an outlier, by its nature, varies significantly from the mean. Therefore, trimming outliers by definition decreases the variance of the sample.

Question 3: Use the average values of t and V (without outliers) to determine the predicted RMSE of Q.

See Table 1, Row 1. For implementation details, see lines 8, 9 of the code provided.

Question 4: Compare the predicted RMSE of Q from question 3 to the correct unbiased standard deviation calculated from the data, and comment.

The predicted standard deviation (Table 1, col 8) is generally higher than the actual standard deviation of the sample (Table 1, col 7). This is because the RSME assumes the “worst case” variability between the samples; in the calculation, dV and dt are always present. However, the real distribution can contain many samples whose measured value is very close to the real mean. This extra variance predicted by RSME causes its variance to be higher.

Question 5: Using the data Vi and ti=1,2,3,4…N and from the first flow rate measurement (fast flow rate, large cylinder), calculate the standard deviations sv and st, and compare them with δV and δt, respectively. (δV and δt are provided in the data sheet). Comment on the difference observed.

st and sv are given in cols 3 and 5 of Table 1. We can see that the time variance is approximately the same for each run. Except for the slow, large run, which had several significant outliers, in time and in volume. Therefore, while the calculated flow sample variance is low (because the outliers in volume and time disappear when those two quantities are divided), the variance for the individual quantities of time and volume variance is high, because of the presence of those outliers.

In general, we can also see that the volumetric sample variance scales with the size, with the larger volume having a large absolute variance in volume – suggesting that the sample variance scales with the container size. We somewhat expect this; for the larger container, we cannot see the exact volume, so we pull the container out of the water stream with more variance than in the smaller container. However, we expect the percentage variance to be roughly equal between the two sized containers.

On flow rate, however, we see the opposite pattern. The slower flow rate has a lower variance. This makes sense; with more time passing, the variance caused by human error of pulling the cylinder out of the flow earlier or later (which causes a variance in the time) is smaller compared to the absolute time that the cylinder is in the flow, while the variance in container measurement remains the same. Therefore, the calculated flow variance is smaller for the lower flow rate.

At Home Experiment (Question 6)

Rather than measuring volume directly, I used the mass of the water to infer the volume (assuming that 1mL = 1g.). My scale can read up to 1g in precision; unfortunately the weight precision is not measured. The timekeeping device was a CASIO W-800H, which could measure to millisecond precision. I estimate that I could keep time within approximately 0.2s of precision. Fig 5 shows the home experiment setup.

I measured a high flowrate of 103 mL/s and a low flowrate of 28.1 mL/s, as shown in the last two rows of Table 1. The variance as calculated was roughly 4 times higher for the high flow rate, which we expect, because the calculated flow rate was approximately 4 times higher. The variance expected by RSME was approximately 4 times higher for the fast flow than for the slow; this suggests that the estimates for time and volume variance of 0.2s and 1mL are roughly accurate.

The distributions of the home experiment are shown in Fig 3 and 4. We see that after removing the outliers using the 3-sigma rule, the distributions are the same. This was in line with experimental observations; for the most part, experimental technique remained similar throughout.

However, this does not mean that the flow rates were un-biased. Because of how the water clung to the vessel when pulling it in and out of the water stream, there was water that was brought to the scale from the outside of the vessel that was not necessarily measured during each sample (visible in the right image of the ensemble of Fig 5.). Thus, it's possible that the water could have made the weight of the vessel erroneously high, and therefore biased every sample uniformly high. However, the difference between the weight of the scale dry in comparison to with that water was measured to be less than 1g; therefore, it's unlikely that it contributed too much to the bias.

Fig 5: Home Experiment Setup. Left: the “low flow” setting; Center: the beaker and timer; Right: the scale mid-experiment with residue from the container.

Tables

	Time mean	Time variance	Volume mean	Volume variance	Flow Rate mean	Flow Rate variance	Flow Rate predicted RSME
In Lab, vol=large, rate=fast	5.05	0.27	684	38.7	135	2.38	13.4
In Lab, vol=large, rate=slow	10.6	0.73	642	45.8	60.2	0.78	2.88
In Lab, vol=small, rate=fast	0.55	0.11	39.6	5.66	75	17.5	65.9
In Lab, vol=small, rate=slow	0.59	0.14	35.8	8.05	62	9.38	52.3
At Home, rate=fast	2.91	0.24	300	24.5	103	1.93	7.08
At Home, rate=slow	7.09	0.24	199	5.97	28.1	0.43	0.81

Table 1: Standard Deviations and RSME Values For the 6 Distributions

Data Analysis/Methods

Data analysis was performed in the python script found at:

https://gist.github.com/rland93/c8d994f2755ded5bd42346c49506dc8b

The charts were generated with seaborn, data manipulation was performed with pandas, and some statistical measures were determined using the scipy.stats module.

Graphs

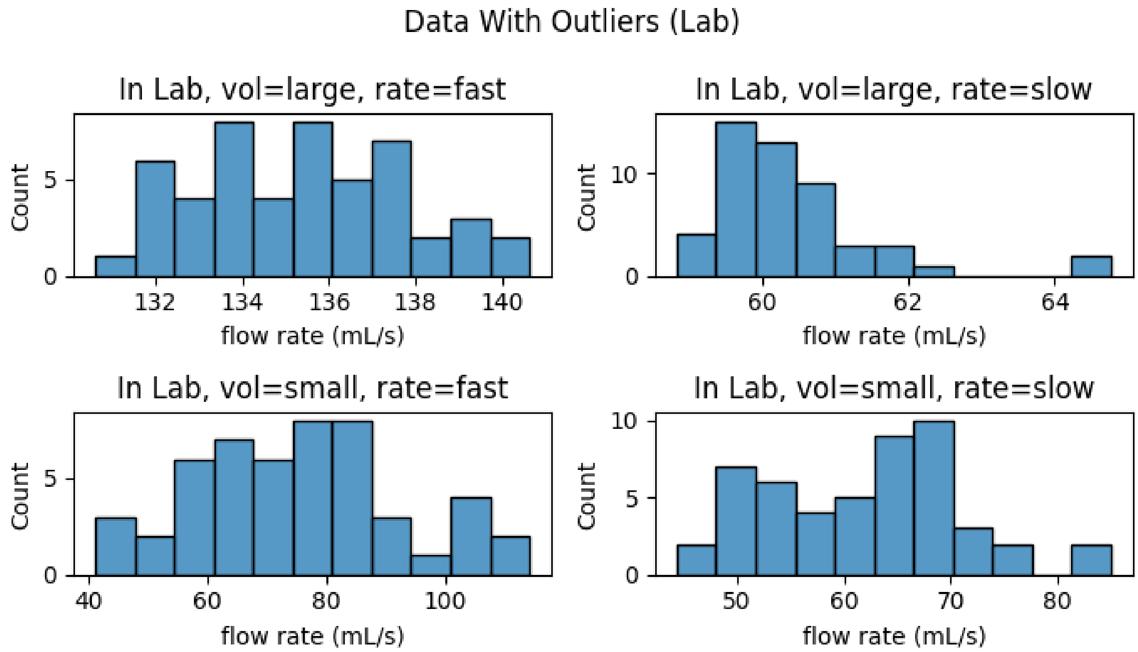

Fig. 1, Lab Data with outliers

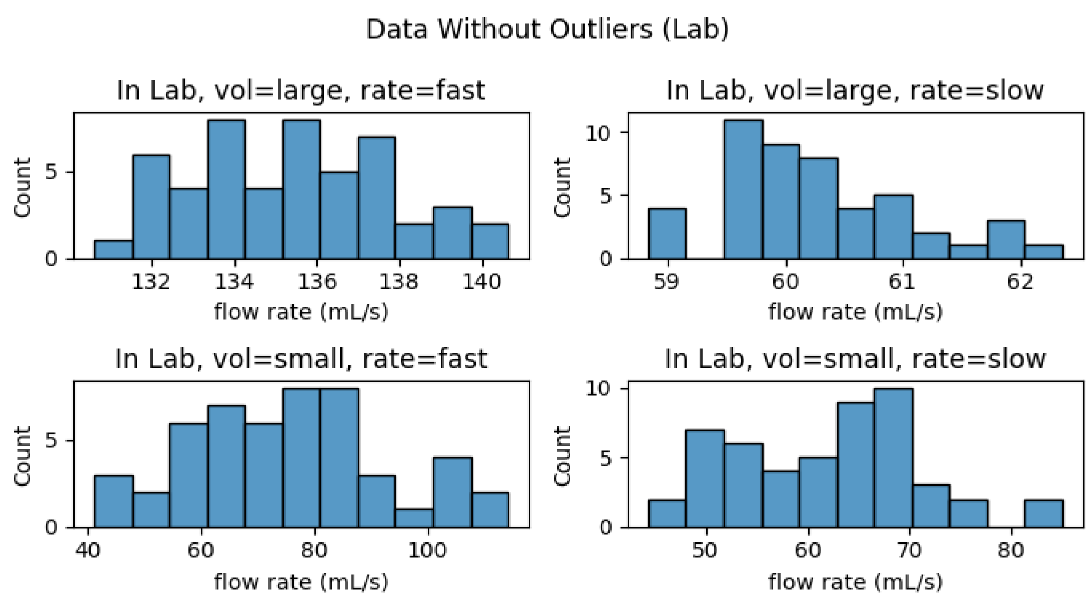

Fig 2, Lab Data without outliers

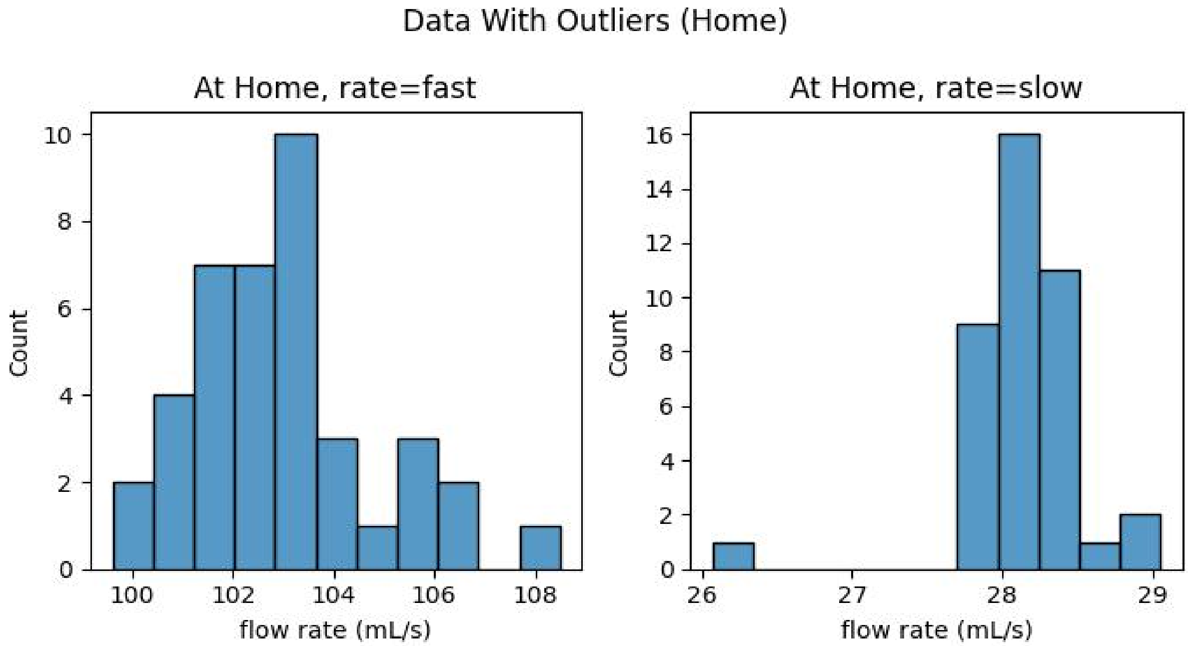

Fig 3, Home Data with outliers

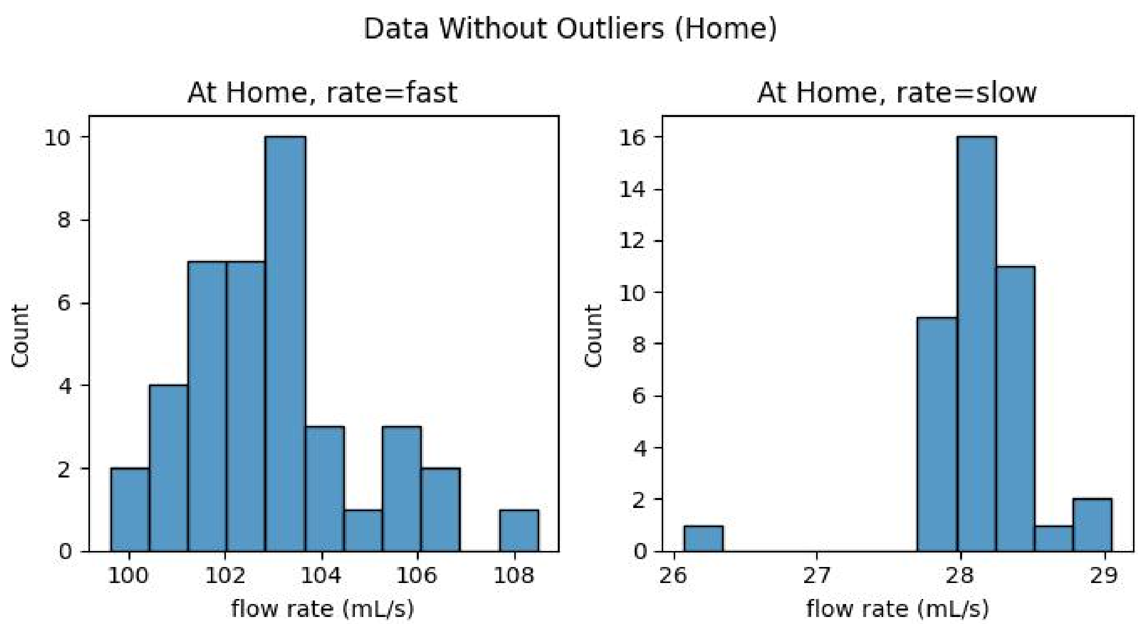

Fig 4, Home Data without outliers