Lab 4 – Warm-Up Worksheet – Estimating measurement uncertainties when there is dispersion in the data
Reference: Taylor, Sections 4.2, 4.3, 4.4

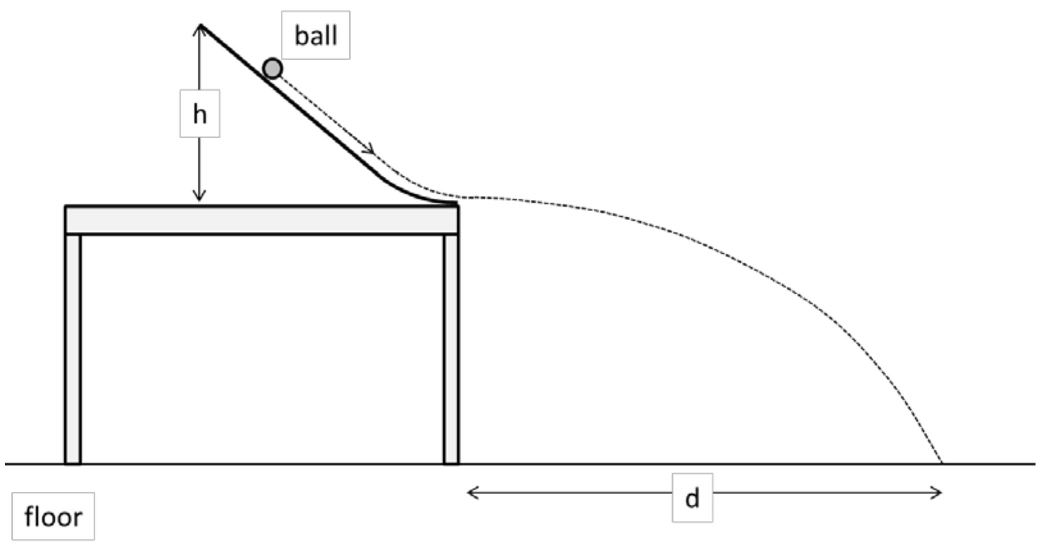
Group: ___________________________

Analyst: ___________________________

Experimenter: ___________________________

Recorder: ___________________________

1. **Arithmetic mean**
A student performs the following experiment in the physics laboratory. A wooden slope is clamped near the edge of a table. A ball is released from a height *h* above the table as shown in the diagram. The ball leaves the slope horizontally and lands on the floor a distance *d* from the edge of the table. Special paper is placed on the floor on which the ball makes a small mark when it lands and the student uses a long ruler to measure d and *h*.



The student rolls the ball from a height *h* = 78.0 mm and measures the distance to the centre of the spot to be *d* = 650.4 mm.

Just to be sure, the student rolls the ball again from the same height, but this time measures *d* = 660.6 mm.

a. What should the student record as the best estimate of the distance *d*?

The best estimate for *d* = ___________________ mm.

How did you choose this value?

____

b. A third measurement yields *d* = 659.1 mm. The best estimate for *d* = ___________________ mm.

c. A fourth measurement yields *d* = 669.6 mm. The best estimate for *d* = _________________ mm.

d. We see that there is clearly **dispersion,** or scatter, in the readings for *d*. Why don't all the data agree?

Actually, it is usually not possible to identify a single reason for what causes the observed scatter in the data. Even if the student does the experiment as carefully as possible, there will still be dispersion in the readings of *d*. How can we deal with this dispersion?

Usually the average value, or **arithmetic mean**, is the best value to use. But the mean is changing as the student does more measurements! How many readings does the student have to take when the data is showing scatter?

To investigate, let's look at the student's data.

The student takes 50 readings - you can find the data in a file in the lab module. Use your spreadsheet program to calculate a **running average**, the average calculated up to and including each reading.

e. You can see that the average jumps around at first, but then settles down. By looking at the table, decide how many rolls from h = 78.0 mm would be reasonable to give a reliable average for *d* and justify your answer:

2. The experimental uncertainty

We have already seen that the best approximation for the value of the measurand (*d* in this case) is given by the average value:

$$\bar{d} = \frac{1}{N}\sum_{i=1}^{N} d_i \ ,$$

where N is the number of readings (in this case 50).

What about the uncertainty in any given measurement? Early in the course we used half of the spread of the data as an estimate of the uncertainty in any given measurement. This method assumes that the data is spread uniformly across this range – in reality the distribution of the data will not be uniform, with more data close to the mean value than at the limits of the range.

What we are looking for is the "average scatter" of the data. One way to get the average scatter is to take each reading d_i, subtracted from the mean, add them all up and divide by the number of readings.
Can you see what would be the result of calculating this? In other words, we could calculate the average scatter by using the formula $\frac{1}{N}\sum_{i=1}^{N}(\bar{d} - d_i)$.

a. Use your spreadsheet to calculate this quantity for this data set:

$\frac{1}{N}\sum_{i=1}^{N}(\bar{d} - d_i) =$ ______________ mm

b. The result of this calculation will always be zero. Why?

Therefore we use the square of the deviations $(\bar{d} - d_i)^2$ and calculate $\frac{1}{N-1}\sum_{i=1}^{N}(\bar{d} - d_i)^2$. This is called the **variance** of the data. Note that we have defined this by dividing by *N-1* and not *N* (this is a subtlety but is usually justified because we are using the mean value in the calculation). The variance gives us a measure of the spread of the data. The square root of the variance is called the **standard deviation** σ,

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(\bar{d}-d_i)^2}.$$

c. Calculate the variance and standard deviation of this set of data:

$$V(d) = \frac{1}{N-1}\sum_{i=1}^{N}(\bar{d}-d_i)^2 = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \text{ mm}^2$$

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(\bar{d}-d_i)^2} = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \text{ mm}$$

Now we have two quantities, the average and the standard deviation, which give us the best approximation of the measurand together with a measure of the spread of the data, respectively.

3. The uncertainty of the mean

The **uncertainty of the mean value** decreases as the number of data points increases and the actual value is given by

$$\sigma_{mean} = \frac{\sigma}{\sqrt{N}}.$$

Logically this makes sense – the more measurements we do the better we should know the mean. The actual relation is a result of propagation of the uncertainties in the calculation of the variance.

Use your spreadsheet to calculate the uncertainty in the mean for this data set:

σ_{mean} = ______________________ mm

4. Summary

a. Write down the second data point with its experimental uncertainty

d_2 = __________________ ± ____________________ mm

b. Write down the mean value with its experimental uncertainty

$\bar{d}$ = __________________ ± ____________________ mm