EXPERIMENT 2
Reaction Time

Objectives

- to make a series of measurements of your reaction time

- to make a histogram, or distribution curve, of your measured reaction times

- to calculate the "average" or mean of these reaction measurements as a "best value"

- to calculate the "standard deviation" or "uncertainty" associated with an individual measurement

- to calculate the "standard deviation" associated with the mean value

- to compare your calculations with the data displayed on the histogram, and with the prediction from the "normal" or Gaussian distribution

- to discuss the significance of data comparison when the spread in values is large

Theory

Two of the main purposes of this experiment are to familiarize you with the taking of experimental data and with the reduction of such data into a useful and quantitative form.

In any experiment, one is concerned with the measurement of some physical quantity. In this particular experiment it will be your reaction time. When you make repeated measurements of a quantity you will find that your measurements are not all the same, but vary over some range of values. As the spread of the measurements increases, the reliability or precision of the measured quantity becomes poorer. If the measured quantity is to be of any use in further work, or to other people, it must be capable of being described in simple terms. One method of picturing measured values of a single quantity is to create a histogram. The histogram is a diagram drawn by dividing the original set of measurements into intervals or "bins" of predetermined size, and counting the number of measurements within each bin. One then plots the frequency (the number of times each value occurs) versus the values themselves. The histogram has the advantage of visually presenting the distribution of readings or measurements. Figure 1 shows a typical histogram for a set of observations. When placing the values into bins, one systematically puts values that occur on the bin limits into the next higher bin.
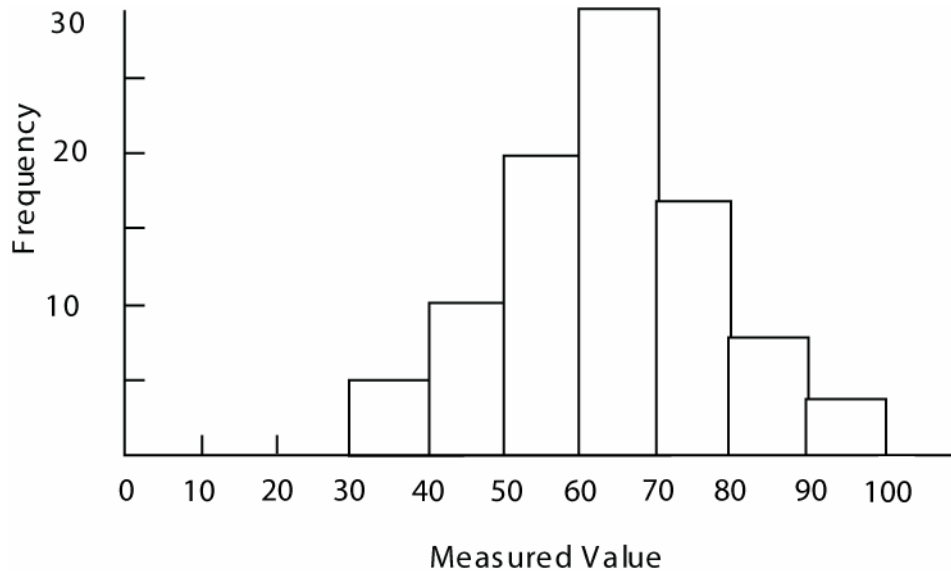
Figure 1 Typical histogram (bin size = 10)

When analyzing data with a histogram, the distribution often times suggests that there is a "best" or most likely value, around which the individual measurements are grouped. From an intuitive approach one might say that the best value is somehow related to the middle of the distribution, while the uncertainty is related to the spread of the distribution. The following formulas, which we will define, will in general only have significance for symmetrical distributions. Using mathematical statistical theory it turns out that the best value is nothing more than the arithmetic average or *mean* of our measurements, which we will denote with the symbol: $\bar{x}$.

$$\text{Best value} = \text{average} = \text{mean} = \bar{x} = \frac{\sum x_i}{N}$$

where

$$\sum x_i = x_1 + x_2 + x_3 + \ldots + x_N$$

N is the total number of measurements
x_i are the values of individual measurements (i.e. x_1, x_2, x_3 etc)

We now need to define a quantity that is connected with the width of the distribution curve. We use a quantity that quantifies the deviations of the individual readings from the central (mean) value of the distribution. This is called "variance" and is defined as follows:

$$\text{Variance} = \frac{\sum (x_i - \bar{x})^2}{N - 1}$$

where

$$\sum (x_i - \bar{x})^2 = (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots$$

The quantity that we will associate with the uncertainty of a single measurement is called the "standard deviation" (denoted by "s") and is simply the square root of the variance.

2

$$s = \sqrt{\mathrm{var}\,iance}$$

We are also interested in the uncertainty of $\bar{x}$. That is, by how much $\bar{x}$, calculated for different sets of data, are likely to deviate from each other. This uncertainty is characterized by s_m, the width of the experimental distribution of values of $\bar{x}$ or *"Standard Deviation of the Mean"* which is calculated by

$$s_m = \frac{s}{\sqrt{N}}$$

Note: the larger the number of measurements made of a quantity by an experimenter the smaller the random uncertainty associated with the mean value.

If the number of readings is very high and the bins are small, the histogram approaches a continuous curve and is called a "distribution curve". Many theoretical distribution curves have been defined and their properties evaluated, but the one that is most significant in the theory of measurement is the Gaussian or "Normal" distribution. If all of the experimental data that you have obtained correspond to one and the same physical quantity, then for very large number of measurements they will be described by the Gaussian distribution with its peak at the average value $\bar{x}$.
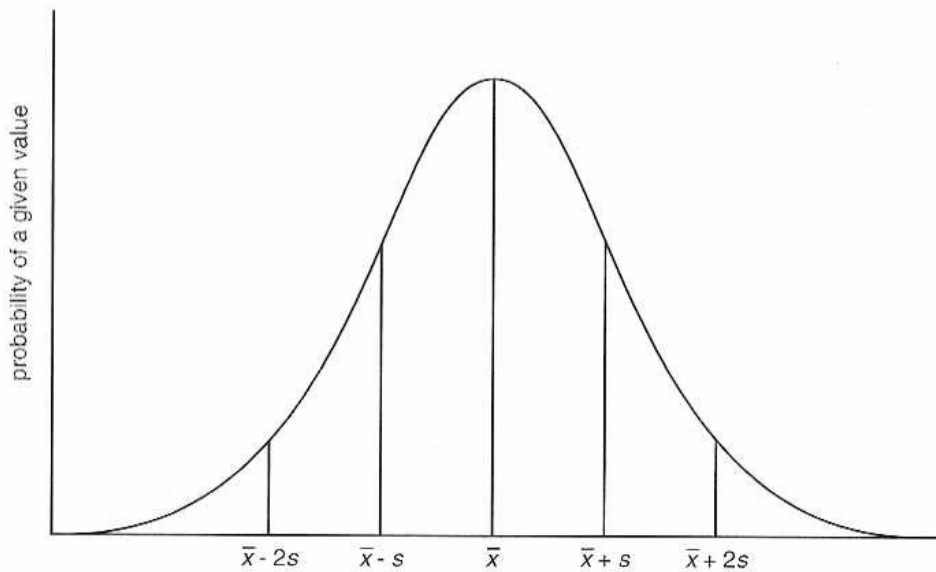

Figure 2   Gaussian distribution curve

Some of the properties of this continuous distribution are that it is symmetric around a peak value and that it falls to zero on either side of the peak, giving it a "bell shaped" appearance (see Figure 2 above). We use the Greek letter "σ" to represent the standard deviation when referring to a Gaussian distribution and "*s*" for the standard deviation calculated from FINITE sets of observations ("*s*" is the best estimate of "σ" for a finite set of observations). When considering Gaussian distributions, the area enclosed by the range ±σ around the peak

3

will contain 68% of the area of the curve (or 68% of the measurements). This means that an individual measurement has a 68% chance of falling within a region $\pm\sigma$ around the peak, or "mean" value, of the distribution. An area bounded by the range $\pm 2\sigma$ will contain 95% of the area of the curve and therefore represent a 95% chance that an individual measurement will fall within this region of the distribution. This is illustrated in Figure 2.

Procedure

One member of the group will "secretly" start the digital timer with a remote switch. The subject will use the "stop" switch to halt the timer as fast as his or her reflexes permit. Record the reaction time and reset the timer for the next run. MAKE AT LEAST 10 PRACTICE RUNS BEFORE RECORDING ANY DATA. Record the 25 measurements of the reaction time in the ".xls" spreadsheet in the *Reaction time* folder, in the "Time" column.

You will first be calculating $\overline{X}, s, s_m$ for ONLY your first five measurements, or N=5. Do these calculations by hand first on a sheet of paper to be handed in with rest of your work at the end of the lab session.   Be sure to show all your work. Now redo the calculations using the spreadsheet and compare the results:

First you will have to calculate the mean value for the first 5 measurements. DO NOT JUST TYPE IN THE VALUE YOU CALCULATED BY HAND. Use the techniques learned in the first lab to enter the formula for the mean for the first 5 values. In order to do this you should use the *Excel* function "SUM" to automatically find the sum of a group of numbers. The basic form the function takes is SUM(First coordinate: Last coordinate). Therefore, typing the equation "=SUM(B12:B16)/5" in the appropriate yellow "Mean for N=5" cell will result in *Excel* calculating the mean of the quantities entered in cells B12, B13, B14, B15 and B16. Press return to "activate" the formula. Compare the *Excel*-calculated mean to the one you figured out by hand. If they don't match up, check your work and your *Excel* formula!

The second and third columns in the spreadsheet, which are grey, are used to help you calculate the standard deviation for N=5. Look at the formula for standard deviation and make sure you understand how these two columns are steps on the way to calculating that quantity. You will type the appropriate formula into the first cell of each column and use your fill-down skills to complete them. NOTE: Since you will be filling down, you can NOT use the cell coordinate of the N=5 mean in your formulas to complete columns **C** and **D**. If you do, the fill-down function will continue to use values down the column from the mean as it completes the fill-down command. *Excel* assumes you want to increment the cell containing the mean value by one for each subsequent calculation unless you tell it otherwise! There are three ways to let Excel know that you do not want this cell incremented as it calculates each new value:

1. You can just type the value of the mean itself into your formula, or
2. You use the procedure from lab 1 to rename the cell, or
3. You can explicitly reference the cell using the "$" command in *Excel*. To explicitly reference cell **C19**, enter "C19" into your formula.

In the last experiment you renamed the cells to avoid this problem. In the experiment, try explicitly referencing the cell containing the mean value. You want the second column to contain $(x_i - \bar{x})$. To do this enter "=B12-C19" in cell C12 and then use the fill down technique to have Excel calculate the value for the other four measurements (in cells C13, C14, C15 and C16). The third column should contain $(x_i - \bar{x})^2$. To calculate this value, enter "=C12^2" in cell D12 and then use the fill down technique to have Excel calculate the other four measurements.

Next, use your values in column **D** to create your formula for the yellow N = 5 standard deviation cell. The formula for the standard deviation is: $s = \sqrt{\text{var}iance} = \sqrt{\dfrac{\sum (x_i - \bar{x})^2}{N-1}}$

Notice the numerator of the ratio under the square root is just the *sum* of the values in column **D** and the denominator is just N-1 = 5-1 = 4. To enter this formula in Excel type "=SQRT(SUM(D12:D16)/4)" into cell C21.

You can now use this cell to calculate the standard deviation of the mean for N = 5. Recall, $s_m = \dfrac{s}{\sqrt{N}}$ so type "=C21/SQRT(5)" into cell C23 to calculate the standard deviation of the mean.

Now repeat the calculations in your spreadsheet to find the mean, s and s_m for N = 10 and N = 25. **You do not need to do any hand calculations for N=10 or N=25!** Don't forget to change for formulae to reflect new number of measurements. Also, as the number of measurements you use changes, so will your values for the mean and standard deviation of your data. In addition, your values for the standard deviation of the mean should get smaller as N gets larger.

Use *Kaleidagraph* to plot a histogram of your 25 measurements. To do this, go to "Gallery", select "Stat" and then choose "Histogram". You should now set your bin sizes on the histogram. The bin size should not be smaller than 25% of your standard deviation value. To set the bin size, go to the "Plot" menu and select "Plot Options". A "Plot Options" window should open. Click on "Histogram", choose "Specifying the Bin Size" and click "OK". Next, go back to the "Plot" menu and select "Axis Options". An "Axis Options" window should open. Click on "Limits", enter your bin size in the space provided and click "OK". Now, make sure your graph is properly labeled. Print out your histogram. On your printout, <u>clearly and neatly</u> mark the values or ranges $\bar{x}$, $\bar{x} \pm s$ and $\bar{x} \pm s_m$ by hand.

Questions

1) Calculate the actual percentage of all your 25 measurements which lie between $X \pm s$ (show these calculations). Does your result differ from those expected for a pure Gaussian distribution?

2) Suppose your lab partner was talking to the students at an adjacent lab table when you started the timer. As a result, the time registered on the timer when it was stopped was 10 seconds. How many standard deviations (s) does this represent? Should you include this data point with the rest of your data? Why or why not?

3) If you have already taken 25 measurements, how many more measurements of reaction time would you have to take to reduce s_m by a factor of two? Justify your response.

4) It is often necessary to compare two different pieces of data or results of two different calculations and determine if they are compatible (or consistent). **In just about every experiment in this course you will be asked if two quantities are compatible or consistent.** The following describes how to determine if two pieces of data are consistent (or compatible). Use this procedure to answer the question at the end and use it as a reference whenever you are asked if two pieces of data are compatible or consistent. Let's denote the pieces of data by d_1 and d_2. If $d_1 = d_2$ or $d_1 - d_2 = 0$, clearly they are compatible. We often use Δ (pronounced "Delta") to denote the difference between two quantities:

$$\Delta = d_1 - d_2 \text{ (1)}$$

This comparison must take into account the uncertainties in the observation of both measurements. So our comparison becomes, "is zero within the uncertainty of the difference Δ?" Which is the same thing as asking if:

$$|\Delta| \leq \delta\Delta \text{ (2)}$$

To perform the comparison, we need to find $\delta\Delta$. The data values are $d_1 \pm \delta d_1$ and $d_2 \pm \delta d_2$. The addition/subtraction rule for uncertainties is:

$$\delta\Delta = \delta d_1 + \delta d_2 \quad \text{(3)}$$

Equation (2) and (3) express in algebra the statement "d_1 and d_2 are compatible if their error bars touch or overlap." The combined length of the error bars is given by (3). $|\Delta|$ is the separation of d_1 and d_2. The error bars will overlap if d_1 and d_2 are separated by less than the combined length of their error bars, which is what (2) says. For more information on using uncertainties to compare data, see section 4 of Appendix A

Two red blood cell counts are $(4.53 \pm 0.14) \times 10^6 \dfrac{cells}{cm^3}$ and $(4.84 \pm 0.28) \times 10^6 \dfrac{cells}{cm^3}$.

Is the probability reasonably good or very bad (choose one) that these measurements are from the same human? Evaluate the difference and comment.

Checklist

Your lab report should include the following four items:

1) the filled spreadsheet
2) the histogram
3) your calculations sheet
4) your answers to the questions