\

**THE BANANA LAB: AN INVESTIGATION INTO THE STARCH CONTENT OF RIPENING BANANAS**

**Created By Bill Martin**

**Introduction by Ted Constantine**

**Introduction**

There are a number of different definitions that adequately describe ripening. Ripening is the fourth of the six major developmental stages that a banana flower ovary passes through. During the ripening stage the flower ovary or fruit undergoes both microscopic and physical changes. These changes include changes in color, texture, and aroma. This process is comparable to an animal going through puberty with the fluctuating hormone levels and rapid physical changes (Goodwyn & Mercer,1972).

The process of ripening is accompanied by a major increase in

respiration called the climacteric rise. The climacteric rise is generally preceded by a rise in ethylene production and increased C02 liberation. There are a number of possible reasons for why respiration increases during the climacteric and the two most plausible are as follows: (A) oxidative phosphorylation is increased by the increase availability of ADP, possibly as a result of an increased requirement for ATP; or (B) the liberation of a natural uncoupler of phosphoryation from respiration, which will allow the latter to proceed at a faster rate (Goodwyn & Mercer 1972).

Ethylene is an endogenous ripening hormone that is

responsible for the stimulation of ripening in the maturing fruit. It is a simple hydrocarbon (H2C=CH2). Despite the fact that it is a gas under physiological conditions of temperature and pressure, ethylene is soluble in the cytoplasm which allows it to be transported throughout the cell (Raven, Evert, Eichhorn). In bananas there is a small amount of ethylene present prior to climacteric and there is a significant increase in production just before the increase in respiration rate. Bananas make a low quantity of ethylene prior to

the climacteric and the removal of ethylene stops ripening completely.

However, at the climacteric, the presence of ethylene acts via a positive feedback mechanism (increases its own production) to stimulate the production of more ethylene, which greatly hastens the ripening process. The role of ethylene in bananas is very important. Bananas become more sensitive to ethylene as they mature. One possible reason for this may be the decrease in concentration of an endogenous inhibitor. Theoretically, when the concentration of this inhibitor drops below a certain threshhold level, the ethylene may act to stimulate the ripening process. Alternatively, ethylene may start to trigger a series of ripening events by binding to a metallic receptor from which it can be replaced competitively by C02 (Goodwyn & Mercer 1972).

The biosynthesis of ethylene begins with the amino acid

methionine. The methionine reacts with ATP to create a compound called S-adenosylmethionine, which is abbreviated SAM. The SAM compound is then split into two different molecules. One of the molecules contains a ring composed of three carbon atoms. This compound is known as 1-aminocyclopropane-1-carboxylic acid or ACC. The ACC is then converted into ethylene, C02 and ammonium by enzymes on the tonoplast. The tonoplast is the membrane surrounding a vacuole in the cell. This conversion from ACC to eythylene is a rate limiting step that slows down the over all speed of the metabolic pathway producing ethylene.

On an intracellular level many things are occurring during

ripening. Many of these changes may result from increase in the permeability of membranes which is increased by ethylene's action. This allows enzymes and substrates, which would normally be separated, to come together. This is an attractive idea in relation to the conversion of chloroplasts into chromoplasts, which is accompanied by the destruction of one set of pigments, the chlorophylls, and a massive synthesis of another, usually carotenoids.

This also includes the conversion of starch to sugar in amyloplasts as starch degrading enzymes, amylases, move from the cytoplasm into the amyloplasts. After ripening occurs, the mitochondria also disintegrate and lose respiratory efficiency, but they do remain intact and functional until the fruit is fully ripe (Goodwyn & Mercer. 1972). This suggests that the ATP produced by respiration in the mitochondria is still needed to drive chemical reactions including the synthesis of proteins for enzymes involved in cellular changes occurring during ripening and whose activity was increased by ethylene's action. The continued requirement for ATP to drive protein synthesis is further supported the fact that while during climacteric the synthesis of rRNA stops altogether, the synthesis of mRNA and thus translation into proteins continues.

During ripening the pectin substances that the cell wall relies on for its integrity undergo changes which affect its consistency. The main changes that the pectins seem to undergo include a shortening of the polymer chain length. This change has important secondary effects on the bonding between pectin and other cell wall constituents such as cellulose and hemicelluloses. In general the cellulose level is altered very slightly, but there does seem to be a reduction in the level of hemicelluloses (Goodwyn & Mercer, 1972). Pectin esterase levels in banana pulp remain constant and total and insoluble pectins decrease during ripening while soluble pectins increase (Nagy & Shaw, 1980). These changes cause cells to slip past each other and account for the fruit softening. This slippage is due to the degrading of the principal component of the middle lamella of the cell wall which is a pectin.

Protein levels in bananas increase during ripening. Ripe fruit contain only low levels of nitrogen, about 50% of which is contained in proteins. Typical protein levels for most fruit are around 40mg/100g fresh wt., but they can reach about 1g/100g in high protein fruit such as bananas. Slight changes in gross protein levels have been reported occasionally during maturation and ripening but nothing spectacular seems to happen. Various changes in the free amino acid pattern in ripening fruit have been reported but no specific pattern has emerged from this data (Goodwyn & Mercer,1972).

Very little is known about the changes which lipids undergo during ripening. Usually fruit contain only low levels of lipids and these have not been seen to change during ripening (Goodwyn & Mercer, 1972). Lipid changes do affect the constituents of the cell membrane and the membrane surrounding the amyloplasts. This might account for the ease of depletion of amyloplasts.

In almost all fruit there is a gradual decrease in starch stores and starch stores sometimes totally disappear at the climacteric. The demands for respiration generally utilize about 20% of the storage carbohydrate, whilst the remainder appears as sugars such as fructose, glucose, and sucrose. The cell wall pectins and hemicellulose may be the source of these additional sugars (Goodwyn & Mercer, 1972).

Many compounds and unique mixtures of compounds give rise to the characteristic flavor and aroma of fruit. There is a high level of tannins in many immature fruit which render them unpalatable because of their bitter taste and astringency. Goodwyn and Mercer (1972) asserted that the virtual disappearance of tannins during ripening of fruit is one of the major factors in making bananas less astringent, whereas Nagy and Shaw (1980) cited that the polymerization of tannins are responsible for the loss of the bananas astringency.

***Special thanks to Ted Constantine for his work in crafting this literature review.***

***Related Articles:***

Physiological and chemical changes during ripening of Costa Rican bananas harvested in different seasons.

<http://journal.ashspublications.org/content/121/6/1157.full.pdf+html>

**THE LAB:**

The central purpose of this lab is to study the complex carbohydrate (starch) changes in ripening bananas. Microscopic analysis of banana cells at three ripeness stages (green, ripe, and overripe) is done to examine changes in the number amyloplasts (starch storing plastids stained by iodine) per cell. Coincidentally, the structure and operation of a compound light microscope, the preparation of a wet mount, the staining of cells in a wet mount, and the observation of the gross morphology of a plant cell is accomplished.

Quantitative analysis of the amyloplast data will include the calculation of means, standard deviation, and standard error of the mean, the graphic representation of the mean data, and inferential statistical analysis of the means by a Student t test to determine significant differences between the means.

**IF TIME ALLOWS:**

Testing can be done for the presence of reducing sugars such as glucose and fructose at the three stages of ripeness. The reducing sugar tests are done using a Benedicts qualitative reducing sugar test. In addition, the presence of polypeptides (proteins) can be tested for using Biuret reagent, and the presence of lipids can be completed using Sudan IV, a nonpolar, fat soluble dye.

**BANANAS AND THE SCIENTIFIC METHOD**

"The scientific method, in its simplest form, is a process by which conclusions are made based on carefully made observations. Central to the scientific method is the formulation of a hypothesis." A hypothesis is a tentative explanation or educated guess for an observed phenomenon. An experimental hypothesis is a statement that is testable and can be answered yes or no by the outcome of a properly designed experiment. The **experimental hypothesis is drafted in an “if… then experiment, results” format** before the experiment is conducted; the experiment is constructed so that its results will hopefully give a clear yes or no answer to the assertion made in the experimental hypothesis, evaluated by the use of inferential statistics. A clearly stated experimental hypothesis tells what should be observed in the experiment if the hypothesis is correct or incorrect, and the results of the experiment either support or falsify (fail to support) the original experimental hypothesis, respectively (Tietjen, 1991).

"A theory is a broad description of a phenomenon supported by many lines of evidence and is held at a higher level than that of a hypothesis in science. Theories often tie together many disciplines within the sciences (such as the theory of evolution, the cell theory, or atomic theory). They serve as generalizations that help to organize scientific thought and create models for visualization. The terms "theory" and "hypothesis" cannot be used interchangeably in the sciences" (Tietjen, 1991).

**NOTE:** A “**null** hypothesis” is a statistical statement that asserts that two means, treatment and control for example, are equal to each other. This is not to be confused with the experimental hypothesis described above; the null hypothesis is used in inferential statistics to support or fail to support the experimental hypothesis.

• Write a **long form** experimental hypothesis predicting the relationship between banana peel color and the average number of amyloplasts per banana cell in the flesh of a banana:

**Format:** If…then experiment briefly, results if hypothesis is correct.

\

• Write a **long form** experimental hypothesis predicting the pattern of ripening within the banana itself (for example, inside to outside or other pattern within the banana fruit flesh) that can be answered using the basic experimental method described in Part I of this lab.

**Format:** If…then experiment briefly, results if hypothesis is correct.

• Use a random number table or generator to collect the amyloplast per cell data to enable you to select cells in which to count amyloplasts in an unbiased manner. This helps to prevent any strong belief in your experimental hypothesis from skewing your data collection such that you find data to support your hypothesis when, in reality, your hypothesis may be incorrect. One must always be on guard against biased data collection if true relationships are to be found.

**PART I**

**MICROSCOPIC ANALYSIS OF BANANA CELLS AND RIPENING CHANGES IN AMYLOPLASTS PER CELL**

1) With a toothpick, take a small smear from the very outer surface of the middle third of a green banana, place it on a glass slide with two drops of water. Mash it gently with the toothpick to make a thin, uniform slurry on the slide. Put a cover slip on it.

• Observe the unstained slide under low power, then high power. **In the space below, draw a few (three or more) of the cells that you see from the slide.** **Use Pencil!** **Label the cell structures/organelles you can see.**

OUTER (green cells, unstained) OUTER (green cells, stained)

• **NEXT:** Add a drop or two of iodine to one edge of the cover slip of the outer surface sample and draw the iodine stain through by placing the edge of a paper towel on the opposite edge of the cover slip. **Draw a few (three or more) of the cells that you see from the slide. Label the cell structures/organelles you can see in them in the space above.**

**Q 1**• Starting at low power and moving to high power, **what do you observe?** **What color are the stained amyloplasts**? **What is this reaction a positive test for?** **Explain the significance of this observation.**

**Q 2** • **Can you see (round) nuclei in the cells? If not, what does this indicate about mRNA transcription from genes (DNA)? If not, why might banana cells lose their nuclei?**

2) Starting with the **stained** **green** **banana outer surface cell sample** on high power, use the random number table or digital generator to select ten different cells whose stained amyloplasts you will count and record in the data table. To select the first cell, start anywhere in the random number table, but continue from left to right, top to bottom from your starting point. If you start with a 5, for example, count the first five cells that you see in the field moving left to right, and count the number of amyloplasts in the fifth cell. You may need to use the fine focus adjustment to change the depth of field and “tease out" the edges of one amyloplast from another, especially in the greener samples. Move the slide to change fields or start counting again from the fifth cell. If the next number in the number table is 3, count the first three cells you see in the new field (left to right and top to bottom), and count the number of amyloplasts in the third cell. Do this for a total of ten cells and record the number of amyloplasts per cell in the data table under green banana, outer surface.

3) **Clean the slide**. **After washing your hands! taste a fresh sample** of the green banana from one of the outer two thirds of the original banana you cut your sample from so you can relate the taste to your observations. **Record your description of the taste** in the data table.

4) Complete steps 1-3 for the ripe and overripe banana samples. Draw your diagrams for each **STAINED** sample in the spaces below and label all visible structures/organelles.

**Ripe Banana Cells- With iodine stain Overripe Banana Cells - With iodine stain**

**DATA TABLE - TASTE TEST**

**Q 3**• Record the results of your taste test; also **relate the results** to what was observed

in the cells using the microscope.

SAMPLE COMMENTS



RIPE

OVERRIPE

**INDIVIDUAL** DATA TABLE: NUMBER OF AMYLOPLASTS PER BANANA CELL FOR OUTER SAMPLES AT THREE STAGES OF RIPENESS

|  |  |
| --- | --- |
| RIPENESS STAGE | OUTER SAMPLE |
| GREEN | 1 . 6.  2. 7.  3. 8.  4. 9.  5. 10. |
| RIPE | 1. 6.  2. 7.  3. 8.  4. 9.  5. 10. |
| OVERRIPE | 1 . 6.  2. 7.  3. 8.  4. 9.  5. 10. |

**USE EXCEL TO CONSTRUCT A SECOND TABLE TO COMPILE THE WHOLE CLASS'S DATA FOR EACH STAGE OF RIPENESS.**

**Q 4**• The collection of data from additional samples is called "replication." **Why is**

**replication of an experiment useful in drawing conclusions?**

**Q 5** Using your personal data (10 cells) and the whole class's data AND AN EXCEL SS CALCULATION determine the average number of amyloplasts per cell for:

**A. The "Outer Sample" means for each stage of ripeness. Record below**.

GREEN class average amyloplasts per cell.

RIPE class average \_\_\_\_\_\_\_\_\_\_ amyloplasts per cell.

OVERRIPE class average \_\_\_\_\_\_\_\_\_\_\_ amyloplasts per cell.

**B. Give the standard deviation (SD) for each class means:**

GREEN \_\_\_\_\_\_\_\_\_\_

RIPE \_\_\_\_\_\_\_\_\_\_

OVERRIPE \_\_\_\_\_\_\_\_\_\_

**C. Give the standard error of the mean (SE) for each class mean:**

GREEN \_\_\_\_\_\_\_\_\_\_

RIPE \_\_\_\_\_\_\_\_\_\_

OVERRIPE \_\_\_\_\_\_\_\_\_\_

**D. What *should* happen to SD and SE as sample size increases? Why use SE instead of SD as a measure of data variation?**

·

\

**Q 5 continued**

**E. GRAPHING AND ANALYZING THE DATA FROM COUNTING AMYLOPLASTS PER**

**BANANA CELL…**

•**Using the CLASS MEANS** prepare a **scatter plot** of the average number of amyloplasts per cell vs stage of ripeness. Add the appropriate trend line, its equation, and R2 value.

For the graph you will need to determine the following and add them to your graph in the appropriate places:

The "Independent Variable" is \_\_\_\_\_\_\_\_

plotted on the horizontal (X) axis.

The "Dependent Variable" is \_\_\_\_\_\_\_\_\_

plotted on the vertical (Y) axis.

Graph TitIe: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Write a **VERY DESCRIPTIVE TITLE** that tells ALL of the information you are conveying with the graph.

**Print the Excel Graph to submit.** Remember to select a unit scale for each axis that takes advantage of most of the paper.

**Q 6** Use spreadsheet formulas (label/identify calculated values clearly on your SS) to calculate the values needed to conduct a Student t test. Compare, two at a time, the **mean** number of amyloplasts per cell for the green, ripe and overripe samples (compare green to overripe cells only as needed).

**A)** Green cell average \_\_\_\_\_\_\_\_\_\_\_\_\_\_ Ripe cell average \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

For the **green vs ripe** show: the calculated student t test value \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

The number of degrees of freedom (df) for your calculation \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

The t table value for your number of df at p=0.05; p=0.01 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

The **statistical** null hypothesis: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Your conclusion about a significant difference or not between the means and WHY:

**B)** Ripe cell average \_\_\_\_\_\_\_\_\_\_\_\_\_ Over ripe cell average \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

For the **ripe vs over ripe** show: the calculated student t test value \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

The number of degrees of freedom (df) for your calculation \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

The t table value for your number of df at p=0.05; p=0.01 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

The **statistical** null hypothesis: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Your conclusion about a significant difference or not between the means and WHY:

C) Green cell average \_\_\_\_\_\_\_\_\_\_\_\_\_\_ Over ripe cell average \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

For the **green vs overripe** show: the calculated student t test value \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

The number of degrees of freedom (df) for your calculation \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

The t table value for your number of df at p=0.05; p=0.01 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

The **statistical** null hypothesis: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Your conclusion about a significant difference or not between the means and WHY:

**Q7** **Providing** the means you compared and the statistical t test calculated values and critical table values from Question 6, justify/prove your conclusion relative to your experimental hypothesis: what is the relationship between banana peel color and the number of amyloplasts per cell as the banana ripens? **Be concise.**

**NOTE:** You are presenting a methodical, reasoned analysis for the accuracy of your initial hypothesis; follow the sequence below to learn the pattern for a good analysis:

1. Begin by restating your hypothesis.
2. Next, state whether or not the results supported, failed to support, or contradicted the hypothesis.
3. Then, bring the averages into methodically making a case for the accuracy of your assertion. Do this by describing what the t test comparisons showed, giving each t test calculated value and the critical value with probability level noted (0.05 or 0.01).
4. Make a closing statement to summarize your key conclusion that you supported with your evidence above.

**Q8 A.** What happens to the amount of starch in a cell’s amyloplasts as the banana ripens from green to overripe?

**B.** What *happens* to the starch if it decreases - *where does it go* *and/or what is it used for*? Base your answer on ***BOTH*** your *taste test* observations/evidence and the background information in the Introduction or similar background information from web source(s) that you find (cite your web source titles and URL’s after your answer).

**Q9** What causes/brings about any changes that are observed in the starch quantity? Specifically, (A) what initiates the process of starch breakdown and then (B) what actually breaks down the starch?

Use the Introduction and/or web research for this information and cite your web source title(s) and URL(s) after your answer.

**PART II**

**IDENTIFYING THE PATTERN OF RIPENING WITHIN A BANANA’S FLESH**

**RECALL FROM EARLIER IN THE LAB:** Write an experimental hypothesis predicting the pattern of ripening within the banana itself (for example, inside to outside or other pattern within the banana fruit flesh) ***that can be tested using a similar, but modified experimental method described in Part I of this lab.***

**Q11** Restate your second, long form hypothesis here predicting the pattern of ripening within the banana itself. **Format:** If…then experiment briefly, results if hypothesis is correct.

**Q12** Create a spreadsheet to compile the data your group has collected.

**Q 13** Using your group’s data AND Excel SS cell formulas, determine the average number of amyloplasts per cell for each of your samples and list clearly the averages using the format in Question 5 (no need to provide standard deviation or standard error of the mean here):

**Q 14**  Using Excel, prepare and **PRINT** **ONE** **scatter plot** of the average number of amyloplasts per cell vs. \_\_\_\_\_\_\_\_(depends on your hypothesis) to most clearly depict your hypothesized pattern of ripening. Label the axes and give the graph a very descriptive title. Add the appropriate trend line(s), each line’s equation, and R2 value, and **provide a key** **and clearly label the identity of each line** if there is more than one line on the graph, even if printed in color!

**Q 15** Use your spreadsheet (label/identify calculated values clearly on the SS) to calculate the values needed to conduct a Student t test to compare the mean number of amyloplasts per cell for the samples you have chosen.

Provide information for each calculation using the previous format from Question 6:

**MODEL ONLY:**

**A)** Green cell average \_\_\_\_\_\_\_\_\_\_\_\_\_\_ Ripe cell average \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

For the **green vs ripe** show: the calculated student t test value \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

The number of degrees of freedom (df) for your calculation \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

The t table value for your number of df at p=0.05 and p=0.01 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

The statistical null hypothesis: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Your conclusion about a significant difference or not between the means and WHY:

**Replicate this model as many times as is needed for all necessary comparisons of two means at a time.**

**Q16**  **A)** **Providing** the means you compared and the statistical t test calculated values and critical table values from Question 15, justify/prove your conclusion relative to your experimental hypothesis about the pattern of ripening in a banana? How does the pattern of ripening seem to proceed in the banana- inside to outside, outside to inside, uniformly, or some other pattern, throughout the flesh? Frame your answer as you did in Question 7. Be concise!

**B)** Based on background information and your research into the way that bananas are treated after they are harvested from a tree, why do you think the ripening pattern proceeds this way?

**Q 17** How could this experiment (from Q 11) be modified to more confidently answer question 16A?

**Q 18** Why does storing bananas in a paper bag cause them to ripen faster?

(Reason it out and then research it to support (or refute) your reasoning and document your source(s) titles and URL’s.)

**Standard Deviation vs Standard Error**

How to calculate SD and SE and their meaning (video): <http://www.wikihow.com/Calculate-Mean,-Standard-Deviation,-and-Standard-Error>

Calculations of the mean, standard deviation, and standard error are most useful for analysis of normally distributed data. One standard deviation about the central tendency (plus and minus relative to the mean) covers approximately 68 percent of the data, 2 standard deviation 95 percent of the data, and 3 standard deviation 99.7 percent of the data. The standard error gets smaller (narrower spread) as the sample size increases.

In short, standard deviation describes variation about the mean in single sample (100 banana cells at one stage of ripeness counted). Standard error of the mean describes the precision of the sample mean, or degree of uncertainty around the estimate of the mean of the whole population (all banana cells at one stage of ripeness).

Below is from: <http://www.protocol-online.org/biology-forums-2/posts/11239.html>

The terms "standard error" and "standard deviation" are often confused.1 The contrast between these two terms reflects the important distinction between data description and inference, one that all researchers should appreciate.   The standard deviation (often SD) is a measure of variability. When we calculate the standard deviation of a sample, we are using it as an estimate of the variability of the population from which the sample was drawn. For data with a normal distribution, about 95% of individuals will have values within 2 standard deviations of the mean, the other 5% being equally scattered above and below these limits. Contrary to popular misconception, the standard deviation is a valid measure of variability regardless of the distribution. About 95% of observations of any distribution usually fall within the 2 standard deviation limits, though those outside may all be at one end. We may choose a different summary statistic, however, when data have a skewed distribution.3   When we calculate the sample mean we are usually interested not in the mean of this particular sample, but in the mean for individuals of this type—in statistical terms, of the population from which the sample comes. We usually collect data in order to generalize from them and so use the sample mean as an estimate of the mean for the whole population. Now the sample mean will vary from sample to sample; the way this variation occurs is described by the "sampling distribution" of the mean. We can estimate how much sample means will vary from the standard deviation of this sampling distribution, which we call the standard error (SE) of the estimate of the mean. As the standard error is a type of standard deviation, confusion is understandable. Another way of considering the standard error is as a measure of the precision of the sample mean.   The standard error of the sample mean depends on both the standard deviation and the sample size, by the simple relation SE = SD/(n)1/2. The standard error falls as the sample size increases, as the extent of chance variation is reduced—this idea underlies the sample size calculation for a controlled trial, for example. By contrast the standard deviation will not tend to change as we increase the size of our sample.   So, if we want to say how widely scattered some measurements are, we use the standard deviation. If we want to indicate the uncertainty around the estimate of the mean measurement, we quote the standard error of the mean. The standard error is most useful as a means of calculating a confidence interval. For a large sample, a 95% confidence interval is obtained as the values 1.96xSE either side of the mean. The standard error is also used to calculate P values in many circumstances.   The principle of a sampling distribution applies to other quantities that we may estimate from a sample, such as a proportion or regression coefficient, and to contrasts between two samples, such as a risk ratio or the difference between two means or proportions. All such quantities have uncertainty due to sampling variation, and for all such estimates a standard error can be calculated to indicate the degree of uncertainty.   In many publications a ± sign is used to join the standard deviation (SD) or standard error (SE) to an observed mean—for example, 69.4±9.3 kg. That notation gives no indication whether the second figure is the standard deviation or the standard error (or indeed something else). A review of 88 articles published in 2002 found that 12 (14%) failed to identify which measure of dispersion was reported (and three failed to report any measure of variability).4 The policy of the BMJ and many other journals is to remove ± signs and request authors to indicate clearly whether the standard deviation or standard error is being quoted. All journals should follow this practice

-Pradeep Iyer-

**Student t test: two tailed, unequal variances.**

ABOUT: Two sample t test, HOW: <http://ccnmtl.columbia.edu/projects/qmss/the_ttest/twosample_ttest.html>

A GREAT, simple t test calculator (unequal variance, BOTH tails/sides): <http://in-silico.net/tools/statistics/ttest>

\*\*\*Another calculator (NOT paired, copy and paste data to n=99): <http://www.physics.csbsju.edu/stats/t-test.html>

t test table of probabilities, use two tailed, 0.05 or 0.01 probability level:

<https://www.easycalculation.com/statistics/t-distribution-critical-value-table.php>

**Excel** can be used to do a t-test: <http://office.microsoft.com/en-us/excel-help/t-test-function-HP010335701.aspx>

Use the equation: T.Test (array1, array 2, 2, 3) means (data set 1, data set 2, 2 tailed test, unequal variance).

From Freshman Physics Foundations by Ben Morgan, Pomfret Science Department Faculty

Accuracy and Precision and Significant Figures \*

There is some degree of uncertainty in every measurement. We usually try to distinguish between two contributions to uncertainty: limitations of accuracy, and limitations of precision.

Accuracy expresses how closely the measurement comes to the true or accepted value. A measurement may be extremely reproducible, giving the same results each time, yet it may not actually measure that which it is supposed to measure. In such a case, the accuracy of a result may be very uncertain. Finding the source of inaccuracy requires making the same measurement by several other methods. It is, therefore, not easy to specify the accuracy of a measurement.

Precision or reproducibility expresses the variation found when experiments are

performed in which the same procedure is used repeatedly. Any measurement has some uncertainty resulting from limitations ofthe measuring device and the experimenter's ability to use the device. Gross errors due to blunders or lack of knowledge of proper procedure for making observations can reduce precision drastically, but careful scientists are not apt to be troubled by these.

Laboratory measurements usually require that you estimate the fraction beyond the smallest division on the instrument. This last figure is, therefore, uncertain but is customarily included in recording the measurement in your data. Those figures about which you are certain and the one more final figure about which there is some uncertainty are called **Significant Figures.**

The uncertainty of a recorded measurement will be assumed to be ± 1 (plus or minus one) in the last significant figure. Thus 34.5m is a measurement with three significant figures, and we would assume the uncertainty to be ± 0.1m. This says that we would expect repeated measurements to yield values no bigger than 34.6m and no smaller than 34.4m. Uncertainty may be indicated in other ways, but we will normally use the "significant figure" method.

It is sometimes difficult to determine whether a zero in a recorded measurement is significant. Of course, the person who reads the instrument knows this, but he must follow accepted rules concerning significant figures if he is to communicate the information accurately to others who may use his data. In order to overcome the difficulties of communication which may arise where zeros are included in the expression of a measured quantity, the following rules are customarily used:

1. All nonzero digits are significant: 126.34 has five significant figures (5sf)

2. All zeros between two nonzero digits are significant: 120.007 has 6sf.

3. Unless specially indicated by the context to be significant, all zeros to the left of an understood decimal point but to the right of a nonzero digit are not significant: 109,000 has only 3sf.

**Significant Figures (cont.)**

4. All zeros to the left of an expressed decimal point but to the right of a nonzero digit are significant: 109,000. now has 6sf.

5. All zeros to the right of a decimal point but to the left of a nonzero digit are

not significant: 0.00467 has 3 sf. (The single zero conventionally placed to the left of the decimal point in such an expression is never significant.)

6. All zeros to the right of a decimal point and to the right of a nonzero digit are significant: 0.04060 and 30.00 both have 4sf.

7. Exact numbers, counted numbers of items (# of amyloplasts per cell) or defined units of measure (ex. 1 kg = 1000 grams) have an infinite number of significant figures.

**Significant Figures in Calculated Results**

When we combine two or more measurements in making various types of calculations, we will be interested in the uncertainty of the result; we will want to know how many significant figures to retain in our answer. The following conventions have been adopted to avoid claiming less or more uncertainty in calculated results than the data justifies. In general, your final result is no more certain than the *least* certain of the measurements used.

1. Addition and Subtraction

For all numbers being added or subtracted, retain only as many figures to the right of the decimal point as are included in the number with the fewest figures to the right of the decimal point.

3.14

45.2

+ 304.60

352.94 should be rounded-off to read 352.9

2. Multiplication and Division

For all numbers derived from multiplication or division, retain the same number of significant figures in your answer as are include in the component with fewest significant figures.

8035

X 3.14

25,229.90 should be rounded-off to 25,200

To say that 3.14 has 3sf is to say that this number might be as high as 3.15 or as low as 3.13. If it were 3.15, then our answer would come out 25310.25. If it were 3.13, then our answer would come out 25,149.55. Thus there is uncertainty about the third figure in our answer, and we have no right to suggest that it is certain by including more figures in our answer beyond it. Therefore, we have written the result as 25,200 which has 3sf.