Microbiological Procedures

Notes for Lecture and Laboratory

Prepared for Micro 3053
by W.H. Lorowitz

Table of Contents

1 WHAT IS "SCIENCE?". . . 1
- *The Meaning of "Science"*. . . 1
- *The Structure of Science* . . . 2
- *How is Science Done?*. . . 4
- *Making Observations*. . . 5
  - Ordered Observations. . . 5
  - Surveys.. . . 5
  - Experiments. . . 6

2 WRITING SCIENCE. . . 7
- *The Laboratory Notebook*.. . . 7
  - Why Keep a Notebook?. . . 7
  - The Proper Tools. . . 7
  - Setting Up the Notebook. . . 8
  - Recording Experiments. . . 8
  - Other Uses. . . 9
  - Electronic Notebooks.. . . 10
- *Preparation of Scientific Manuscripts and Experimental Lab Reports*. . . 10
  - Additional Information. . . 12
  - Quality Standards Checklist. . . 13

3 MOCK WOUND. . . 15
- *Background*. . . 15
- *Materials*.. . . 15
- *Procedure*. . . 16
- *Results*.. . . 17
- *Conclusion*. . . 17

4 MOCK STOOL.. . . 18
- *Background*. . . 18
  - Biochemical Tests. . . 18
    - Triple Sugar Iron (TSI) Agar. . . 18
    - SIM Agar. . . 19
    - Urea Agar. . . 19
    - Methyl Red-Voges Proskauer (MR-VP) Broth. . . 19
    - Simmons Citrate Agar. . . 20
  - Media. . . 20
    - MacConkey Agar. . . 20
    - Eosin Methylene Blue (EMB) Agar. . . 20
    - GN Broth, Hajna.. . . 20
    - Salmonella-Shigella (SS) Agar.. . . 20
- *Materials*.. . . 21
- *Procedure*. . . 21
- *Results*.. . . 23
- *Conclusion*. . . 23

5 *STATISTICS*. . . 24
What Are Statistics and Why Do We Use Them?. . . 24
Descriptive Statistics. . . 24
Measurement Scales. . . 25
Central Tendency. . . 25
Variability. . . 26
Graphs. . . 27
Histograms. . . 27
Frequency Polygons. . . 27
Box Plots. . . 28
Scatter Plots and Line Graphs. . . 28
Hypothesis Testing. . . 28
Parametric Tests. . . 28
The Student t Test . . . 28
Analysis of Variance. . . 29
Nonparametric Tests. . . 30
Correlations and Regressions. . . 30

6 *USING A DISK DIFFUSION ASSAY TO INTRODUCE STATISTICAL METHODS*. . . 31
Background. . . 31
Purpose. . . 32
Materials (per 4 students). . . 32
Procedure. . . 33
Results. . . 35

7 *MOCK ENZYME KINETICS*. . . 37
Background. . . 37
Competitive Enzyme Inhibition. . . 40
Mock Enzyme Lab. . . 41
PROCEDURE. . . 41
Part I Enzyme Activity and Substrate Concentration. . . 41
Part II: Competitive Inhibition. . . 41
RESULTS. . . 42
DISCUSSION. . . 42

8 *ENUMERATION*. . . 43
Direct Counts. . . 43
Total Counts. . . 43
Viable Counts. . . 43
Indirect Counts. . . 43
Total Counts. . . 43
Viable Counts. . . 44
Dilutions. . . 44
Optical Density Measurements. . . 46
Background. . . 46
Purpose. . . 47

Procedure . . . 47
Direct Microscopic Counts with a Hemocytometer . . . 47
Background. . . . 47
Purpose . . . 47
Materials and Methods . . . 48
Results . . . 48
Viable Counts of Saccharomyces cerevisiae. . . . 48
Background. . . . 48
Purpose.. . . . 49
Materials and Methods. . . . 49
Results . . . 49
MPN Enumeration of Mesophiles and Thermophiles in Soil. . . . 50
PROCEDURE . . . 51
MPN EXAMPLES.. . . . 51
VALUES FOR 3-TUBE MPN.. . . . 52
SOIL MESOPHILES AND THERMOPHILES WORKSHEET.. . . . 54

9 PIPETTING AND SPECTROPHOTOMETRY . . . 55
Background. . . . 55
STATISTICS BACKGROUND. . . . 55
Measures of Central Tendency. . . . 56
The Central Limit Theorem. . . . 56
Hypothesis Testing. . . . 57
Materials. . . . 58
Procedure. . . . 58
Determination of Variability in Pipetting. . . . 58
Determining the Relationship Between Cell Density and Absorbance. . . . 58
Results. . . . 58
Conclusions. . . . 59
Pipetting and Spectrophotometry Lab Worksheet. . . . 60

10 STATISTICS WITH A HEMACYTOMETER. . . . 62
Measures of Central Tendency. . . . 62
Overview. . . . 62
Objectives. . . . 62
Procedure. . . . 62
Hypothesis Testing with Data from Hemacytometer Counts. . . . 66
Objective. . . . 66
Entering and Formatting Data. . . . 66

APPENDIX. . . . 71

Appendix 1. Colony Morphology. . . . 72

Appendix 2. Using Microsoft® Excel to Plot and Analyze Kinetic Data. . . . 73

Appendix 3. Using Microsoft® Excel to Analyze Data from the Disk Diffusion Assay. . . . 85

Appendix 4. Looking for Relationships 93

1 WHAT IS "SCIENCE?"

The Meaning of "Science"

The term *science* has come to mean several things in popular usage. For example, many students are "good at science," usually meaning that they can memorize and understand the information presented in science classes. Another common usage is where *science* is used to describe the general activities performed at so-called scientific institutions. Some may claim that *science* is the search for truth. Although these views touch on things scientific, a proper definition of *science* is that **science is an investigative method for understanding the world around us by evaluating evidence through reason**. Evidence is collected by observing phenomena that may occur naturally or through experimental procedures or surveys. The evidence is analyzed within a systematic and logical framework, generally termed the *scientific method*. So, science is much more about *thinking* than it is about memorizing.

In the scientific method, initial observations are used as the basis for making testable explanations or *hypotheses*. The validity of a hypothesis is tested with additional observations. If the new observations do not agree with the hypothesis, it is rejected and a new hyptohesis is developed. If the new observations support the hypothesis, it is accepted, but it is always susceptible to additional testing. These steps are summarized in Figure 1.1.

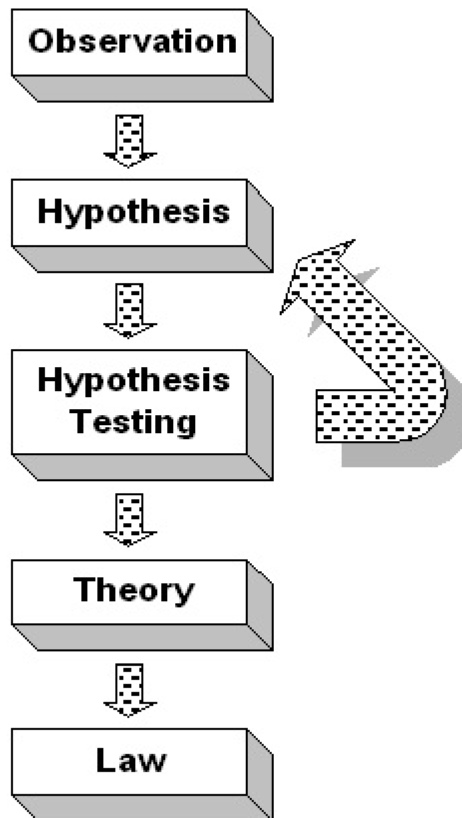


Figure 1.1. Steps in the scientific method.

Recognize that although hypotheses can never be proven, they can be disproved. It may sound contradictory, but scientific progress is made by disproving false hypotheses.

Hypotheses that stand the test of time are usually considered *theories*. In addition to theories, which provide explanations of phenomena, there are also *laws*, which summarize (verbally or mathematically) a large number of individual facts.

The Structure of Science

The logical approach to understanding nature, the *scientific method*, has its roots in ancient Greece. Aristotle invented the formal concept of deductive logic where consequences can be predicted from generalizations; the foundation of deductive logic is to draw conclusions based on premises. The classic example of this:

All men are mortal.	*(premise 1)*
Socrates is a man.	*(premise 2)*
Socrates is mortal.	*(conclusion)*

Notice that the truth of the premises is not considered. They are assumed to be true, with the conclusion following these assumptions. This was the earliest form of logic and arose in opposition to myth. Myths are useful fictions used to provide explanations while maintaining established social orders and recognizing the mystery of the universe. Logic allows truth to be demonstrated and debated.

Aristotle's approach was to draw specific conclusions from general observations. This was challenged by Francis Bacon in the seventeenth century. Bacon introduced *inductive logic*, where general rules are inferred from specific observations. That is, assumptions about a whole group is based on observing a few of its members. For example, we infer taxonomic traits for all white pines based on observations of a few individuals. Induction is the basis of the premises listed above to illustrate deductive logic. Are all men mortal? More people are alive today than existed in the past. That means that the majority of people who were born are still alive! So, why do we believe that all men are mortal? The problem with inductive logic is that there is really no justification for believing the conclusions. However, induction is something we have learned to rely on as a part of everyday life and in science it continues to be a useful tool.

At the same time Bacon was thinking about logic, Galileo developed the idea of a *controlled experiment*. In an experiment designed to study a specific phenomenon all of the variables except the one being investigated are controlled. For example, to investigate the effect of a fertilizer on your rose garden, a simple experiment would be to mix fertilizer into the soil and wait to see if more roses are produced than in the past. But, how would one know if a difference was due to the fertilizer and not to something else, like the weather? A better experiment, where the effect of weather is controlled, would be to fertilize one portion of the rose garden and to leave another portion unamended. Since both portions are exposed to the same weather, it was removed as a variable. Any difference in the yield of roses between the two plots could not be due to the weather and therefore, might support fertilizer as the cause. Of course, other variables, such as watering, would have to be controlled as well.

Karl Popper introduced the concept of *falsification*. Popper believed that nature is objective and that accurate details could be determined through science. Falsification relies on constructing a hypothesis and then looking for evidence to falsify it. The hypothesis that withstands falsification attempts best is accepted. Related to this is *Occam's Razor*, which suggests that "it is wasteful to do with more that which can be done with less." In other words, the simplest explanation is the one to choose.

Isaac Newton introduced the *mathematical model,* where a physical phenomenon is described in terms of mathematical relationships. Mathematical terminology is a very precise language. A good model helps summarize and describe phenomena and is an aid to making predictions. Models can have tremendous value, but their construction requires large amounts of data, determination of the key variables, and elucidation of the relationships between those variables. The equation that describes bacterial growth in exponential phase, $N = N_o \cdot 2^n$, is a straightforward model of what to expect during a particular growth mode. On the other hand, a model describing weather patterns contains many more variables, the relationships of which are not always clear, sometimes resulting in inaccurate predictions. A good place to observe modeling in action is a racetrack; finding people trying to predict future performance using variables from prior races is easy.

In addition to using the scientific method, the scientific community also places social requirements on scientists. A peer review system evaluates the merit and validity of scientific investigations according to the following criteria:

(1) Scientific results should be original. Although it does not hurt to corroborate hypotheses, only novel results are worth reporting on their own. (Most lab classes perform *exercises* rather than experiments, so that specific results can be anticipated. Exercises are most useful for learning new techniques.)

(2) The scientist should be free of bias regarding the outcome of the problems investigated. The validity of the data should be the main concern, not whether or not they support a pet hypothesis. In other words, no motivation should be operating other than the desire to understand a phenomenon.

(3) Results and conclusions should stand on their own merit, without relying on social, ethnic, religious, or personal factors to gain acceptance.

(4) Scientific statements are not to be accepted on faith, but are to be subjected to close scrutiny and falsification attempts. This is why the scientist must play her own Devil's advocate, trying to disprove hypotheses before going public.

(5) Scientific knowledge, including methodology, should be freely available to the public. It is critical that observations can be replicated by anyone with the necessary technical expertise.

(6) The language used to describe all aspects of research must be precise and unambiguous. In science, a report is not subject to interpretation, but must be accepted or rejected solely on the information it presents. ("Say what you mean

and mean what you say.")

Interestingly, even when all of the conditions outlined above are met, introducing information that contradicts existing paradigms is difficult. For example, in the late 1970's when Carl Woese determined that rRNA sequences suggested three evolutionary lines (two prokaryotic, the Bacteria and Archaea, and one eukaryotic, the Eukarya), there was a tremendous amount of resistance from established scientists *who had not evaluated the data* but had learned and accepted Whitaker's five Kingdom scheme. There followed a period of rigorous testing in many labs, all lending support to Woese's conclusions. Today, the three Domain scheme, based on evolutionary relatedness of organisms, is considered correct. (Of course, technological refinements and increasing phylogenetic information continue to "fine tune" the original scheme.) The reaction that Woese faced is common. It is not unusual for a controversial hypothesis to be ignored by the established scientific community. Very often acceptance is not based on the strength of the data, but also requires a shift in paradigms arising from a change in the scientific order, frequently through the demise of older scientists.

How is Science Done?

Science is done by testing hypotheses. The concept of a hypothesis is often unclear, so before proceeding, let's define a hypothesis as a testable explanation for an observed phenomenon. Many times people will define a hypothesis as a prediction or an educated guess, which are not really incorrect, only less precise, in regard to science, than our definition. Predictions play an important part in hypothesis testing, but do not explain a phenomenon. Science requires very precise, unambiguous language, as will be discussed later.

The scientific method requires postulating and testing hypotheses. Hypotheses are tested using deductive logic (if . . ., then . . .) applied to predictions. This approach assumes that there is order in the universe. That is, identical conditions will produce identical results. An important point to recognize is that true predictions may arise from true or false hypotheses. For example, consider the joke where a frog is told to jump. After jumping, one limb is removed and the frog jumps only half the distance. Removal of another limb halves the distance again. Removal of the third limb halves the previous distance. Finally, with all limbs removed, the frog doesn't jump at all, no matter how loudly the scientist yells. Hypothesis: Without legs, a frog becomes deaf. Prediction: No amount of noise will make a deaf (limbless) frog jump. Therefore, it is essential (not just important) to try and falsify a hypothesis and determine if alternative hypotheses exist that explain the data better.

Scientific advancement depends on *rejecting* (not accepting) hypotheses. Acceptance of a hypothesis is always conditional, subject to being refuted at some later time. So, **nothing can ever be proven**. Good scientific research may solve a particularly important problem or produce results from which generalizations can be made.

In any investigation (testing of hypotheses), it is important to control the influence of the factors not being examined. That is, it is desirable that any effects observed be due to the treatment being tested.

If a hypothesis cannot be tested, the scientific method cannot be employed and the hypothesis lies outside the realm of science. Such a hypothesis is not necessarily false, only nonscientific.

Making Observations

Observation is one of the foundations of science. The first step in any scientific investigation is to make an observation or series of observations that can be examined for an explanation (Figure 1.1). There are three ways that observations are made:

1) ordered observations,

2) surveys,

3) experimentation.

Selection of the appropriate strategy or combination of strategies depends on the information sought and the system being studied.

Ordered Observations

Ordered observations are the most common type of observation used by field scientists. Consider a rainforest. An alert observer has no problem finding many examples of species diversity or symbiotic relationships. In fact, the amount of information available is so overwhelming, the observer is forced to ask precise questions and narrow observations to those that bear upon the specific questions. That is the fundamental nature of ordered observations. Not only does the scientist try to answer questions in a specific order, but he also tries to avoid interfering with what is being observed. In other words, it is impossible to study natural behaviors or relationships by placing the subjects into unnatural situations.

Surveys

Unlike ordered observations, surveys require interactions between the observer (surveyor) and the subjects being studied. Surveys are common in, although not limited to, social sciences. The results of surveys can be seen on television news programs everyday. However, surveys are not limited to polls.

There are three basic forms of survey:

1) retrospective

2) current

3) prospective.

A retrospective survey gathers information from the past, for example, historical data or fossil records. The biggest problem with retrospective surveys is that data must be accepted as is. That means that information can be confounded by uncontrolled factors and that older information can be sketchy and less reliable. Consider constructing a family tree. The information used is limited to what’s available (documents, personal recollections), which could be biased and

contain gaps.

A current survey investigates something as it is now, for example, student performance within a class. Although the information is still subject to uncontrolled factors, it is usually easier to get complete information (no gaps).

A prospective survey follows the progress of a selected subject into the future. These are common in nutritional studies, for example the effect of fiber on colon cancer or the effect of wine on heart disease.

Experiments

In an experiment, the observer sets up a situation so that a phenomenon can be observed without being obscured by extraneous factors. This is the most common type of observation used in the laboratory. Well-constructed experiments allow for the unambiguous observation of specific phenomena, allowing the observer to examine cause and effect. Construction of a good experiment is not a trivial skill and will be explored in detail later this semester.

2 WRITING SCIENCE

The Laboratory Notebook

Why Keep a Notebook?

One of the most important tasks a scientist has is to keep an accurate record of experiments. However, maintaining a clear, complete record is easily and commonly neglected. Lab work (except for cleaning glassware) is a lot more fun to do than it is to write about. Unfortunately, experiments that aren't recorded properly are almost useless.

Good records are important to two people: you and anyone else who reads your lab notebook. For you, good records mean you won't have to rely on your memory to recall details and could allow you to see patterns in the data. For someone else, good records document your work, allowing it to be repeated and providing support for patents and other credit claims.

A well-kept laboratory notebook saves the researcher time and usually provides the foundation for additional investigations. Faulty memories can lead to repeating the same experiment needlessly, expensive in terms of time and money. Furthermore, data can be misremembered, allowing the investigator to form unsupported hypotheses. Good records allow the scientist (or anyone else) to understand exactly what was done and what was observed, regardless of how long ago the experiment was conducted. An added benefit is that by following a series of experiments, an investigator can often extend the project, asking better, more precise questions. That is, future experiments can be conceived by building on previous knowledge contained in the lab notebook.

The Proper Tools

The value of the lab notebook cannot be overemphasized. As such, it is essential that the notebook itself be sturdy. Pages can be lost too easily from a looseleaf or spiral notebook, or from a notebook with a weak binding Use a string-bound notebook with pages that are strongly and permanently bound together. The cover should be strong to protect the pages and, when necessary, to provide a stiff writing surface. Pages should be numbered. Although some people prefer unlined pages, horizontal or quad rules are more common. A quality notebook will use acid-free paper to prevent yellowing and flaking. The most expensive notebooks also have coated pages to protect against accidental spills. The coating doesn't work well with pencil, but that shouldn't be a concern since records should be kept with indelible ink.

Notebooks are permanent records and must be written with ink. Ballpoint pens are the preferred writing instrument. Their inks are composed of concentrated dyes in organic solvents, usually alcohols. The solvents evaporate almost immediately, leaving sharp lines of fairly permanent pigments. Inks from ballpoint pens are resistant to water and non-polar organic solvents. However, alcohols, such as methanol, can cause the ink to bleed.

Rollerballs, felt-tip pens, and fountain pens use water-based inks. As you might expect, they spread easily if they get wet. Another disadvantage is that the slower evaporation of the solvent (water) allows the ink to diffuse through the paper, so the writing tends to bleed and smudge. Some inks are more light-sensitive than others. Overall, black inks are most permanent; red inks are least permanent.

Sometimes it is necessary to include attachments, for example, photographs, stripcharts, photocopies, or notes recorded away from the notebook. Archival tape that is acid-free and permanent is the best solution, but it is expensive and can be difficult to find. Staples work for short periods but can work loose or tear pages. There are a number of glues or pastes available. They must be permanent and free from chemicals that could degrade paper. Photocopies and laser printer outputs are unstable and can leach onto surrounding papers. They should not come into contact with plastics, including adhesive tape.

Setting Up the Notebook

Professional notebooks start with a signout page that serves as a record of its issue and possession. The next few pages include printed instructions covering legal issues and instructions for making entries. A blank table of contents with spaces for date, title, and page numbers precedes the pages where experiments are recorded. All notebooks should have a table of contents that is filled in as experiments are recorded. Experiments are recorded consecutively in chronological order with no missing or blank pages. In the case where a new experiment is started before another one is completed:

a. Start recording the new experiment on the next blank page.

b. Data from the uncompleted experiment are recorded beginning on the first blank page following the record of the new experiment.

c. The broken records are linked by noting "continued on page . . ." or "continued from page . . ."

Although it is usually neglected, a preface that describes the author and purpose of the notebook is often useful. Typical information includes who is involved in the project (author, supervisors, assistants), the goal of the work, a summary of progress to date, location of the project, and funding. This information can help future researchers and provide support for legal ownership of ideas and inventions.

Ownership is also supported by having each page signed after completion by the author and an uninvolved individual who understands the material. This is a good idea in academic research and is absolutely required in industrial research.

Recording Experiments

The goal in recording an experiment is to be clear, concise, and complete. Clarity requires thinking about things before committing to a permanent record. Concise writing also requires forethought. The idea is to present a straightforward record that is easily understood. Complete records require that information be recorded while it is fresh. If a mistake is made, cross it out with a single line, initial it, and provide a brief explanation. The entry is left legible in case it turns out to be right. Speculation or annotations can be added at the end of the record.

The format for notebook records is essentially the same as that used for a laboratory report or manuscript:

Title: The topic of the experiment. Every experiment should be given a title that clearly

expresses the topic. This provides an easy way to find a specific experiment or its continuation.

Date: The date the experiment began should be recorded in a prominent location, usually at the top of the first page alongside the title. Without a date, it is difficult, if not impossible, to claim first observation.

Purpose: What the experiment is designed to investigate. A statement of purpose is included so that the reader, as well as you, can unambiguously understand what the experiment was designed to investigate. At the very least, the purpose will be described in a single sentence. However, a detailed introduction is clarifies your thinking and makes it easier to write a report. Typically, an introduction provides background information that can help jog your memory or provide the foundation for a reader to understand the experiment. It includes previous knowledge and states an existing problem. The statement of purpose or experimental objective is how you hope to solve that problem.

Materials and Methods: What supplies are needed for the experiment and how they are to be used. A description of materials and methods should provide enough information to allow you or any other capable scientist to repeat the experiment. The list of materials should include suppliers and lot numbers. Methods should be described in detail or properly referenced. This section is easiest to read if it's divided into subsections (*e.g.*, organisms, media, analytical procedures, calculations). ***Tip:*** Diagram the experiment in a flow chart. Seeing the individual steps will help you estimate the time required and the supplies you'll consume.

Results: Observations made during the experiment. Raw data and observations are recorded in a results section. Data should be recorded as it's observed. It's best to set up data tables ahead of time and to fill them in as the experiment proceeds. Observations (*e.g.,* colony morphology, odors, medium color) can be listed; unlike manuscripts, it is not necessary to explain the results in paragraph forms. The primary responsibility is to record the raw data, clearly and completely, but it is a good idea to summarize the data in tables or graphs, similar to how they would be presented in a finished report. This greatly aids manuscript preparation and saves time when going through old experiments.

Discussion: Interpretation of the results. Interpretations of the data can be recorded as a list of conclusions, but it is almost always more beneficial to write a detailed discussion that illustrates your thinking. In either case, the conclusions should address the purpose. It's also appropriate to record conjectures or hypotheses. If possible, additional experiments can be described.

References: Literature cited in the previous sections. Reference citations should be complete, as they would be used in a manuscript

Other Uses

In addition to recording experiments, notebooks can also be used to record literature surveys or ideas. While it's probably best to use note cards, a separate notebook, or a computer database for

lengthy literature reviews, your research notebook is appropriate for notes on a few articles directly related to your project. Notebooks that are witnessed and signed help substantiate ownership and time of origin of an idea. There are numerous accounts of notebooks being used to help determine credit and patent rights.

Electronic Notebooks

Custom software, spreadsheets, and word processors can all be used to maintain laboratory records. In actual practice, it's often easier to use software that can perform calculations and plot data than to record data in a notebook, especially since computers are interfaced to lots of lab equipment. However, there are many drawbacks to using an electronic notebook. First and foremost, computer records alone are not accepted for patents or other legal positions; data can be too easily altered or falsified. Furthermore, electronic databases are not secure and several copies can coexist. Worst of all, many people have poor computer habits and can lose a complete record with a single mishap. The best solution is to use computer applications as necessary, but keep a written notebook and back up data frequently.

Preparation of Scientific Manuscripts and Experimental Lab Reports

The purpose of making a presentation is because you have something worthwhile to convey. Since the essence of communication is based on what the other person perceives and not what you say, it is critical that you express yourself in a clear, efficient manner designed to simplify the reader's ability to understand your meaning. The three fundamental aspects for good communication (including science writing) are:

- plan your structure
- consider your audience
- choose your words

Many publications and journals have style guides for the preparation of manuscripts. Although some differences may occur between style guides, the format of a lab report or research manuscript is fairly specific and is made up of the following sections:

- Title
- Abstract
- Introduction
- Materials and Methods
- Results
- Discussion
- Literature Cited

Subheadings are common within some of the sections, especially the Materials and Methods and the Results sections. Subheadings should help organize information into specific categories that make it easier for the reader to find. For example, Materials and Methods may have subheadings for source of organisms, media, stains, biochemical tests, and statistical analysis. The Results section may have different subheadings, for example cell and colony morphology, biochemical tests, nutritional requirements, growth rates, and DNA homology.

Type the section headings, such as **Introduction**, in boldface, centering them on a line separate from the text body. Subheadings should be in **boldface**, followed by a period, and placed at the beginning of the relevant section.

Science writing is lean and precise. Every sentence has a purpose and each word should be necessary. Metaphor, analogy, hyperbole, and other literary devices are rare. Good grammar and style are essential with such a spare, focused document.

Develop your professional writing skills by comparing your writing with published manuscripts. Pay particular attention to language, style, and organization. Professors, admissions boards, and potential employers will compare your work with others, most especially in regard to how well you demonstrate the clarity of your thinking.

Title: The title should be clear, concise, and unambiguous.

Abstract: The Abstract is a brief summary of what you found with some detail on how the investigation was performed. State the principle objective(s), provide general descriptions of the methods, summarize the results (use data; if "significant," give p-values) and state the principle conclusions. In general, background information is kept to a minimum and tables and figures are not included. Try to summarize each relevant section in just a few sentences.

Introduction: The Introduction contains three areas of information. First, background information is provided to set the stage, furnishing readers with information that allows them to appreciate the investigation and to pique their interest. Think of this as sort of a sales pitch for your work, where you're trying to get the reader interested and thinking about things along a specific line of reason. The writing should be clear, appropriate references should be cited, and meaningless phrases should be avoided (*e.g.,* "Studying microorganisms is very important.") Second, is a statement of the problem that is the focus of the paper, including a synopsis of previous studies and the rationale for your study. The problem statement usually contains the hypothesis of the study and should also indicate why the investigation was worth doing. Finally, the Introduction should include a statement about the purpose of the investigation, which would be to address some aspect of the problem statement.

Keep in mind that a good introduction will

- present the nature and scope of the problem
- acquaint the reader with pertinent literature
- describe the key methods

- state the principle results and conclusions.

Common mistakes made when writing introductions are to try and include a complete review of the literature (leave that for the Discussion) or to go the opposite way and use a lot of empty filler, that is, sentences that do not add anything but words.

Materials and Methods: The information in this section should allow someone competent in microbiology to duplicate the study. Step by step procedures are not given unless they are original. For example, a competent worker does not need to have the procedure for a gram stain explained. If previously published methods are used, a complete citation should be provided along with any modifications used in the current study. Make sure to provide descriptions of how things were done and what equipment and supplies were used. Descriptions of test outcomes and their interpretations can be included. In most cases a flow chart can clarify the sequence of experimental procedures, although they tend to be used only in lab reports, posters, and oral presentations but not in journal manuscripts.

Results: Observations should be described in a narrative. It is almost as if a description of observations was being given over the telephone. The narrative should be supported (not replaced) with summary instruments, such as tables, graphs, and diagrams. Tables and graphs are particularly useful with quantitative data. The same data should not be expressed in a table and a graph. Do not present raw data. Avoid interpreting the results in this section.

Discussion: Interpret the results. What do they mean? Discuss the validity of the observations and techniques. Is there internal consistency between the results? Are the results consistent with external reports? Specific points should be supported by mention of the key results responsible for the conclusion but the goal is to discuss the results, not reiterate them. It is legitimate to make speculations and implications based on your results and the external literature. The first paragraph usually points out the contribution your study makes to your research area, summarizing current literature. It is typical for the Discussion section to end with a paragraph that summarizes the key points established by your research. These conclusions should be related to your introduction, purpose, and hypothesis.

Literature cited: Throughout the report, the sources for key items of information from external sources should be cited. A common way to do this is to follow a statement with the reference number in parentheses. Note that without specific citations, a reference list is a bibliography. In science writing, the references are cited within the text.

Additional Information

Figures and Tables: Figures and tables are used to summarize data described in the text. They should be clear and contribute to making your presentation more easily understood. Figures and tables should be accompanied by a caption that includes a number, title, and brief description of the information being presented. If the figure or table is from an external source, cite the reference and include the complete reference in the Literature Cited section. In the natural sciences it is common practice to place the caption below figures and above tables. The figures and tables must be referred to within the text, usually within parentheses following a statement (*e.g.,* Generation time was 24 minutes (Fig. 3)).

Pay attention to the proper usage of terms:

1. The proper names of bacteria include the genus and the species. Genus is capitalized, species are lower case, and both are italicized or underlined. For example, *Escherichia coli*.

2. "Bacterium" is singular, "bacteria" are plural. ("datum" = singular, "data" = plural; "hypothesis" = singular, "hypotheses" = plural)

3. Use past tense to describe the work you did, present tense for information from an external source.

4. Use scientific notation for large numbers.

5. The scientific way to write dates is DD/MM/YYYY.

6. Only proper nouns get capitalized.

7. Expletives are words used to fill a vacancy (*i.e.,* empty filler) and should be avoided. That is, remove words that are unnecessary. For example, sentences that begin with "It is," "There is," or "There are," can be written without these empty words. Instead of "It is apparent that the results suggest . . .," write "The results suggest that . . ."

8. Write dynamically. Although science writing tends to be in the passive voice ("It has been observed . . ." rather than "It was observed . . ."), the active voice makes reading more interesting. Furthermore, do not use nominalization (changing a verb to a noun). Instead of "we reached the conclusion that . . ." where "reach" becomes the action, write "we concluded that . . ." so that "concluding" is the action.

9. Eschew obfuscation by using concise phrasing and precise wording. Avoid jargon and uncommon words, if possible.

Quality Standards Checklist

☐ **Typing.** All manuscripts and lab reports should be typed. Use a font size of 12 points in a standard typeface, such as Times Roman, for normal text. Margins should be one inch on all sides. Double space the text to make commenting easier; references may be single spaced.

☐ **Neatness.** The manuscript should not be crumpled, torn, or damaged in any other way. Pages should be fastened together with a staple or paper clip. Torn and folded corners do not look professional. Plastic report covers are not necessary.

☐ **Proofreading.** Do not submit manuscripts with spelling errors or poor grammar. Word processors contain spell checkers and grammar checkers. Use them. Proofread the paper before you leave the computer; handwritten corrections are not professional.

☐ **Graphics.** Tables, figures, and graphs should be of high quality. Computer-generated graphics are preferred but hand-drawn graphics are acceptable if they are done with a

high degree of professionalism (*i.e.,* straight lines, even printing, regular proportions). Graphics do not need to be inserted inside the text but may be included on separate sheets attached to the end of the manuscript.

- ☐ **Plagiarism.** Do not submit someone else's work as your own. Use citations with full references to give credit where it is due.

- ☐ **Honesty.** Include all relevant results. Do not "massage" data to create an erroneous interpretation. By the same token, avoid presenting irrelevant information.

3 MOCK WOUND

Background

The goal of this exercise is to isolate and identify the organism(s) considered to be the cause of a wound infection. For this purpose each student is provided with a bacterial mix simulating material from the infected wound. Gram positive cocci such as *Staphylococcus aureus* and *Streptococcus pyogenes,* as well as the gram negative commensals of the intestinal tract, are the most common organisms which cause wound infections. This exercise will be concerned with isolation and identification of the gram positive cocci.

Staphylococcus is a genus of gram positive cocci which appear singly, in pairs, or in irregular clusters. Two species, *S. aureus* and *S. epidermidis,* are frequently found on normal human skin. *S. aureus* is an opportunistic pathogen and can cause abscesses and pustules. *S. epidermidis* is considered nonpathogenic. They can be differentiated from each other very easily with the coagulase test or by observing whether mannitol is fermented. *S. aureus* is positive for both tests, *S. epidermidis* is negative for both.

Streptococcus that are responsible for wound abscesses usually form long chains. Differentiation between *Staphylococcus* and *Streptococcus* is straightforward by testing for catalase activity or, in most cases, growth on Mannitol Salts Agar (MSA). *Streptococcus* does not display catalase activity, *Staphylococcus* has catalase. The 7.5% NaCl in MSA inhibits the growth of most *Streptococcus* but not *Staphylococcus. Enterococcus*, a genus that used to be grouped as a *Streptococcus*, is able to grow with high salt concentrations and will ferment weakly on MSA.

There are four groups of *Streptococcus*, A, B, C, and D. The most common streptococcal group found in human wounds are those in Group A. Members of this group can lyse red blood cells (hemolysis) in three different patterns on blood agar plates. Alpha hemolysis is a partial lysis, demonstrated by a greenish, indistinct zone around the colony. Beta hemolysis describes complete lysis of the red blood cells, producing a clear halo around the colony. Gamma hemolysis is actually the absence of hemolysis; no halo forms around the colony. Beta-hemolytic streptococci in Group A can be differentiated from other beta-hemolytic streptococci by their sensitivity to bacitracin. Although some alpha-hemolytic bacteria are sensitive to bacitracin, the only beta-hemolytic streptococci sensitive to bacitracin are those in Group A.

Materials

Bacterial mixtures (3 species/tube)
- *Staphylococcus aureus* (β-hemolytic)
- *Staphylococcus epidermidis* (γ)
- *Streptococcus pyogenes* (β)
- *Enterococcus faecalis* (γ)

Blood Agar plates
Tryptic Soy Agar plates
Mannitol Salt Agar (MSA) plates

Bacitracin disks
Plasma for coagulase test

Procedure

Pre-lab

- ☐ Use your text to review information on *Staphylococcus* and *Streptococcus* (including *Enterococcus*).
- ☐ Review procedures for making quadrant streak plates and gram stains.
- ☐ Prepare your notebook by entering as much information as possible.
- ☐ Create a flow diagram illustrating the work to be done and the decision-making process.

Day 1

1. Each pair will receive a culture containing three different microorganisms.
2. Each person streaks one Blood Agar Plate for isolation. Each person does a gram stain of the mix.
3. Each table will inoculate Blood Agar Plates in a small cross pattern of known cultures for reference. Gram stain knowns. (What is the importance of this step?)
4. Incubate plates at 37° C for 24 hr.

Day 2

1. Record colony morphologies (See *Appendix 1*), hemolysis patterns, and gram reactions.
2. Streak one of each colony type on TSA plates. (Why?)

Day 3

1. Test each isolate for catalase activity.
   a. transfer a small amount of cells from a TSA plate to a clean microscope slide
   b. apply a drop of fresh 3% hydrogen peroxide (H_2O_2) to the cells
   c. note the production of bubbles (oxygen) that indicates the presence of catalase
2. Innoculate one MSA plate with catalase-positive isolates and catalase-positive reference cultures. (Why only the catalase-positive isolates?)
3. Perform a bound (slide) coagulase test on the catalase-positive isolates and the *Staphylococcus* reference cultures.
   a. add a small amount of cells from the TSA plate to a drop of water on a slide
   b. add a loopful of coagulase plasma to the cell suspension and mix for 20 seconds
   c. formation of clumps is a positive result for coagulase (*S. aureus* is usually positive for coagulase activity and *S. epidermidis* is usually negative)
4. Perform a bacitracin sensitivity test on any isolate that is a gram positive, catalase negative, beta hemolytic coccus. (What organism should this be?)
   1. make a normal quadrant streak of the isolate on Blood Agar
   2. aseptically press a bacitracin antibiotic disk onto the first two quadrants
   3. incubate at 37° C for 24 hr

Day 4

1. Observe and record data from MSA plates. Remember, MSA is a selective and a differential medium. The 7.5% NaCl favors the growth of *Staphylococcus* (although

some strains of *Streptococcus* can grow at high salt concentrations). *S. aureus* can ferment mannitol, producing acid that turns phenol red in the medium from red (neutral pH) to yellow (acidic pH). *S. epidermidis* does not ferment mannitol.

2. Observe and record data from the Blood Agar plates with bacitracin disks. If cells grew adjacent to the disk they are not sensitive to bacitracin and the organism is probably not a beta hemolytic *Streptococcus* in Group A. However, if there is a halo of no growth around the bacitracin disk, the organism is sensitive to the antibiotic and is most likely a Group A *Streptococcus.*

Results

Record your observations at each stage of this exercise. Would a table or plot help summarize any of the data?

Conclusion

Interpret your data. What do the results suggest in regard to the identities of your unknowns? Explain your thought process. Would diagrams help explain your reasoning? Are you confident of your conclusions? Would other tests help?

4 MOCK STOOL

Background

Bacteria in the gastrointestinal tract carry out a wide variety of metabolic reactions, including degradation of macromolecules, vitamin synthesis, and absorption of bile acids. A balanced microbial ecosystem also inhibits the colonization and growth of pathogenic bacteria. As food passes through the gastrointestinal tract, water is removed from digested material, gradually concentrating it as it's converted to feces. Bacteria make up approximately one-third to one-half of the weight of fecal matter. Solid retention time in the gastrointestinal tract is about 24 hours, so bacteria in the lumen double about one or twice each day.

Bacterial counts of stomach contents are low due to the high acid condition (pH2), although the lining is often heavily colonized by acid-tolerant lactobacilli and staphylococci. The small intestine is divided into two parts, with the duodenum resembling the stomach in terms of microbiology. The ileum is less acidic and has a richer population of bacteria, with concentrations of 10^5 - 10^7 per gram.

Most gastrointestinal bacteria are present in the large intestine. The normal flora are predominantly gram positive, obligate anaerobes, *e.g. Bacteroides, Peptostreptococcus,* and clostridia, with counts of 10^{10} - 10^{11} cells per gram of contents. Another component of the population are the enterobacteria, gram negative, facultative anaerobes, *e.g. Escherichia coli, Klebsiella, Enterobacter.* They are present at much lower concentrations, generally less than 10^7 per gram. However, some of the Enterobacteriaceae are pathogenic and can be isolated from the blood, urine, or fecal material of infected individuals. The best known pathogens within this group are *Salmonella* and *Shigella*. *Salmonella* are the most common agent of gastroenteritis in the United States. The main symptoms of *Salmonella* food poisoning are watery diarrhea, sometimes bloody, accompanied with low grade fever. Similar symptoms occur with *Shigella* infections. The presence of pathogens within this group has led to the development of specific media and biochemical tests to differentiate pathogenic organisms from the normal gut flora.

Often the concentration of pathogenic enteric bacteria are very low. For this reason, coliform bacteria (*E. coli, Klebsiella*) are used as indicators of fecal contamination in water supplies. Bacteria are enumerated using MPN dilution series or membrane filter concentration with medium that is selective and differential for lactose-fermenters. In water environments, coliforms have survival rates similar to *Salmonella* and *Shigella*; a large number of coliforms is likely to indicate fecal contamination and water that is unsafe for drinking. In this exercise, the isolation and cultivation scheme employs an enrichment step to increase the proportion of pathogens in the population.

Biochemical Tests

Triple Sugar Iron (TSI) Agar

TSI is used to differentiate members of the *Enterobacteriaceae* based on their ability to ferment

glucose, lactose, and/or sucrose, and to reduce sulfur to hydrogen sulfide. The medium contains phenol red as the pH indicator; sulfur is included as ferrous sulfate and sodium thiosulfate. The medium contains a small concentration of glucose, which all of the *Enterobacteriaceae* can ferment, and 10X greater concentrations of lactose and sucrose. After 24 hours at 37ºC, fermentation of only the glucose will cause the butt of the slant to turn yellow. (The slant will remain red.) If the lactose and sucrose are fermented, both the butt and slant will turn yellow. Gas production will produce bubbles in the butt and may actually split the medium. Any hydrogen sulfide produced will react with iron in the medium to form ferrous sulfide, a black precipitate. Tubes are inoculated by stabbing almost to the base with the inoculating needle and drawing the needle across the slant on the way out. Results must be read within 18-24 hours of incubation at 37ºC to be accurate.

SIM Agar

SIM agar is used to differentiate members of the *Enterobacteriaceae* on the basis of motility, sulfide production, and indole production. Indole is produced as a byproduct of tryptophan catabolism (Figure 5.1). Tubes of medium are inoculated by stabbing 3/4 to the base with an inoculating needle. Check results after incubating at 37ºC for about 24 hours. Sulfide production (from cysteine hydrolysis) appears as a black precipitate (FeS) in the agar. Motility is demonstrated by turbidity extending outwards from the stab but may be difficult to observe due to sulfide production. Read the results again after 48 hours because decarboxylases are inducible are are only produced in an acidic environment in the presence of their substrate. Indole production is determined after 48 hours incubation by adding a few drops of Kovacs Reagent to the tube. A pink color is a positive result for indole

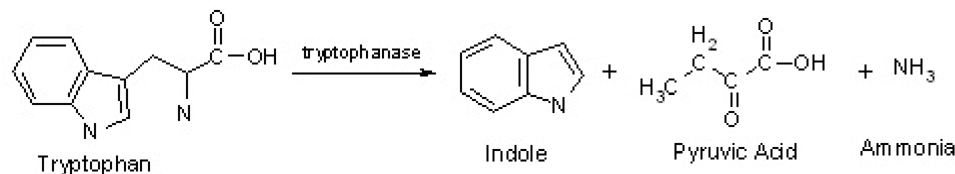


Figure 5.1. Indole production.

Urea Agar

Urea media contain urea and phenol red. Hydrolysis of urea to ammonia and carbon dioxide by the enzyme urease causes an increase in pH, causing the medium to change from yellow to pink. Medium is inoculated by streaking across the surface of the slant. Interpretation of this test should be performed 24 hours after incubating at 37ºC. A pink color throughout the medium indicates rapid activity, typical of *Proteus, Morganella,* and *Providencia*. Pink color only in the slant indicates slower expression of urease activity, typical of *Citrobacter, Enterobacter,* and *Klebsiella*.

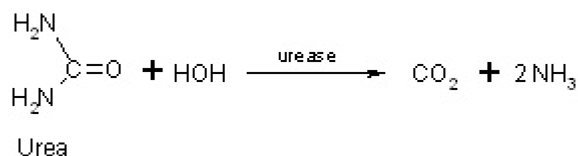


Figure 5.2. Urease activity

Methyl Red-Voges Proskauer (MR-VP) Broth

This test is used to differentiate the between enteric bacteria based on the products from glucose fermentation. Some bacteria, for example *E. coli*, produce acidic end products ($pH < 4.4$) and will cause methyl red to change from yellow (neutral) to red (acidic). Other bacteria within this group, for example *Enterobacter*, produce neutral end products including acetoin. In the Voges Proskauer portion of the test, acetoin is oxidized by KOH in the presence of oxygen to diacetyl. In the presences of α-naphthol, diacetyl forms an intense red color. This test is conducted by inoculating two tubes of MR-VP broth and after incubating at least 24 hours at 37ºC adding methyl red to one tube and KOH plus α-naphthol to the other.

Simmons Citrate Agar

The ability to utilize citrate is useful for differentiating members of the *Enterobacteriaceae*. Organisms that can utilize citrate (Figure 5.3) increase the pH of the medium. The pH indicator, Bromthymol blue, changes color from green (neutral) to blue (alkaline). Tubes are inoculated by streaking across the surface of the slant. Results are read after incubation at 37ºC for at least 24 hours.

Media

MacConkey Agar

MacConkey agar is a selective and differential medium used to isolate and differentiate gram negative enteric bacteria that can ferment lactose (coliforms) from those that cannot ferment lactose. The medium contains crystal violet and bile salts that inhibit the growth of gram positive organisms. Neutral red, a pH indicator, is colorless above pH 6.8 and red below that pH. Thus, coliforms (*e.g., E. coli, Enterobacter*) will form red colonies while organisms unable to ferment lactose (*e.g., Salmonella, Shigella, Proteus*) will be a pale pink, the color of the medium.

Eosin Methylene Blue (EMB) Agar

EMB agar is another selective and differential medium used to isolate and differentiate members of the *Enterobacteriaceae.* Growth of gram positive bacteria is inhibited by methylene blue and eosin. Acid produced from lactose fermentation turns precipitates these dies, producing colonies that are metallic green (typical for *E. coli*) or black. Colonies will be pink if only a small amount of acid is produced. Nonfermenting enteric bacteria form colonies that are colorless or the color of the medium.

GN Broth, Hajna

This is an enrichment medium for enteric bacteria, especially *Salmonella* and *Shigella*. This is a fairly rich medium but it contains sodium citrate and sodium desoxycholate which inhibit the growth of gram positive bacteria and coliforms. The relatively high concentration of mannitol, compared to glucose, favors the growth of mannitol-fermenters like *Salmonella* and *Shigella* over the growth of mannitol non-fermenters like *Proteus.*

Salmonella-Shigella (SS) Agar

SS agar is a selective and differential medium used to isolate and differentiate enteric bacteria that can ferment lactose from those that cannot. Brilliant green dye and bile salts inhibit the growth of gram positive bacteria. The pH indicator neutral red is used to differentiate colonies of lactose fermenters (pink color) from colonies of lactose nonfermenters (colorless). Furthermore, thiosulfate in the medium can be reduced to sulfide by *Salmonella*, showing up as

the black precipitate ferrous sulfide in the center of colonies. (*Proteus* also produces sulfide but ferments lactose, albeit slowly.) There are other media available which are recommended for isolation and differentiation of *Shigella* (Hektoen and XLD agars).

Materials

Escherichia coli
Enterobacter aerogenes
Proteus mirabilis
Salmonella enteritidis
Shigella sonnei
Bacillus cereus
Staphylococcus aureus
MacConkey agar plates
EMB agar plates
GN broth tubes
Salmonella-Shigella agar plates
Triple Sugar Iron (TSI) slants
Sulfide Indole Motility (SIM) agar deeps
Methyl Red-Voges Proskauer (MR-VP) broth
Citrate slants
Urea slants

Procedure

Note: Plates should be interpreted after 24 h incubation then stored in a refrigerator.

Day1

1. Each pair will receive a mixture of four unknown organisms. Record the number of your culture. Each person should make one quadrant streak of the mixture on a MacConkey agar plate and another on an EMB plate.

2. Each person should inoculate a tube of GN broth with the unknown mixture.

3. In groups of four, perform test streaks of the known organisms on MacConkey and EMB agar, dividing the 7 control cultures between two plates of each medium.

4. Each person must make a gram stain of the mixture. Each pair is responsible for performing gram stains with each of the knowns. Record reactions and morphologies of each specimen.

5. Incubate cultures at 37°C overnight.

Day 2

1. Record your observations of colonies on the MacConkey and EMB plates, comparing those from the unknown mix with the control organisms. Select three colonies with different morphologies (there should only be three different morphologies – why?).

2. Streak the colonies onto EMB or MacConkey agar plates to get pure cultures. (It might be a good idea to pick from one type of plate and streak on the other.) Save your control plates in the refrigerator or inoculate fresh plates. Incubate overnight at 37° C.

3. Perform a quadrant streak on an SS plate and a MacConkey **or** EMB plate using the GN broth for inocula. Incubate overnight at 37° C.

Day 3

1. Record your observations of the isolates growing on the different media. Compare the appearance of each colony against the known colonies on the same medium. Record a presumptive identification. Select three different isolates from the same type of medium, if possible. **If you aren't sure you have three different isolates, restreak them on the same medium before moving on to the next step.** (You may want to use the MacConkey plate because it will inhibit swarming by *Proteus*.) The SS plate and MacConkey (or EMB) plate inoculated from the GN broth may be the best source colonies of *Salmonella* or *Shigella.*

2. Once you are sure you have three different organisms isolated (today or tomorrow), streak each of them on Tryptic Soy Agar (TSA) plates as another purification step. After incubation at 37° C, plates may be stored in the refrigerator.

3. Conduct a gram stain on each isolate from the unknown mixed culture.

Day 4

1. Each pair should inoculate three 6-tube setups with the three colonies isolated from the mixture, one isolate per setup. For inocula, touch the top of a colony with an inoculating needle. Each 6-tube setup contains a TSI slant, a SIM agar deep, two MR-VP broths, a citrate slant, and a urea slant.

2. In a group of four, inoculate 6-tube setups with the known organisms.

3. Incubate all cultures at 37°C.

Day 5

1. Record results from media inoculated on Day 4.

2. Based on your results, identify the four organisms that were in your unknown mixture.

Results

Generate your own data table in the format of Table 5.1.

Table 5.1. Metabolic Reactions of Selected Bacteria.

Organism	Gram Reaction	Lactose fermented	TSI slant/ butt	Gas	H_2S	Urease	Motility	I	M	Vi	C
Escherichia coli	–	+	A/A	w	–	–	+	+	+	–	–
Enterobacter aerogenes	–	+	A/A	+	–	–	+	–	–	+	+
Salmonella entritidis	–	–	K/A	w	+	–	+	–	+	–	+
Shigella sonnei	–	–	K/A	–	–	–	–	–	w	–	–
Proteus mirabilis	–	–	K/A	w	+	+	+	–	w	–	d
Bacillus subtilis	+										

Key: – = negative, + = positive, K = alkaline, A= acid, w = weak positive, d = delayed positive
I = indole test, M= methyl red, Vi= Voges Proskauer, C = citrate

Conclusion

Interpret your results. Identify the four different species of bacteria in your unknown mix, **elucidating the logic of your process**. A flow diagram of your decision-making process would be a good summary to accompany your narrative.

5 STATISTICS

What Are Statistics and Why Do We Use Them?

Statistics is a branch of mathematics that is used to analyze incomplete information in order to decrease the level of uncertainty about a conclusion. In other words, statistics allows us to make a decision based on a *sample* of a *population* and to have a specific level of confidence that our decision is correct. Since samples are only a subset of all possible observations, it is critical to evaluate accuracy. Furthermore, there is always some variation between observations and these differences need to be examined to determine if they are real or an artifact of the observation method. Statistical methods are designed to analyze data to help address these difficulties, to provide additional information, and to determine the limitation of the results. Specifically, statistics allow investigators to:

1) use sample sets to make inferences about population parameters,
2) reduce large amounts of data to comprehensible levels summarized by statistical descriptors,
3) determine the level of confidence that supports a particular conclusion.

The use of statistics for data analysis is greatly simplified by selecting the proper tests when designing the experiment.

The two major areas of statistics are *descriptive* and *inferential*. In descriptive statistics, data are collected, organized, summarized, and presented. We focus on summarizing the data in a manner that allows us to determine the type of data we are dealing with and to see any underlying patterns that may exist. The two basic methods available are numerical and graphical. Numerical summaries, like the *mean*, *mode, median,* or *standard deviation* can give us a precise, objective description. Graphical methods allow us to construct a plot, for example a *histogram*, *leaf and stem plot*, or *box plot*. Patterns in the data are more obvious with graphical summaries than with numerical summaries.

Inferential statistics allow us to draw a conclusion about a population from a sample of that population. For example, consider if students who took MICRO 2054 last semester needed an average of four hours to complete the *Identibacter interacticus* assignment and students who took MICRO 2054 prior to last semester required an average of seven hours. Is the difference real or due to chance? Is the magnitude of the time difference accurate or would it change with a larger sample (another class)? These are the types of questions that could be addressed with inferential statistics. The two main methods used in inferential statistics are *estimation* and *hypothesis testing*. In estimation, the data are used to estimate a *parameter* (*e.g.,* the mean) and to construct a *confidence interval* around the parameter. In hypothesis testing, a *null hypothesis* (hypothesis of no difference) is stated and the data are examined to see if the null hypothesis can be rejected.

Descriptive Statistics

Data can be presented in a number of ways, including tables, graphs, and numerical summaries. The appropriateness of a presentation method is determined for each situation, depending on the purpose and the kind of data being examined.

Measurement Scales

Characteristics that are measurable can be classified in one of four measurement scales: the *nominal scale*, the *ordinal scale*, the *interval scale*, or the *ratio scale*. The nominal scale classifies objects or events into mutually exclusive categories. No intermediate conditions can exist, so the object is either "equal to" or "not equal to" the conditions for each category. For example, we might be testing different media with one particular organism and recording "growth" or "no growth." All trials will fall into only one category and an intermediate state is impossible.

The ordinal scale separates objects into categories, like the nominal scale, but in addition, it permits a ranking of "greater than" or "less than" to be performed. The ranking is qualitative, so it is not possible to determine how much two objects differ. An example of an ordinal scale would be a poll of student satisfaction with a class. Although a student who is "very satisfied" is easy to distinguish from one who is "not satisfied at all," it is impossible to provide an actual measurement of their difference. Typically, ordinal data is scored from lowest to highest. In the case of bacterial growth, comparative values might be: 0 = no growth, 1 = some growth, 2 = good growth, and 3 = dense growth.

The interval scale and ratio scale both use defined units of measurement. Not only are qualitative differences noted, but the magnitude of the difference can be measured. Values can be *discrete* and have integer values (*e.g.,* the number of women who became pregnant last month) or *continuous* (*e.g.,* time). To continue with the example of bacterial growth, measurements of actual cell concentrations would use the interval or ratio scale. The main difference between the two scales is that the ratio scale sets zero equal to zero, while the interval scale uses an arbitrary value for zero. Temperature uses an interval scale, with the freezing point of water used to arbitrarily set 0° C. The ratio scale is used for more tangible physical properties, like mass or time. Since most of the variables we are measuring have continuous and discrete values, we usually use the ratio or interval scales.

Central Tendency

If two die are rolled, the frequency of the values obtained approaches a distribution that can be predicted from the number of combinations possible. That is, there is only one way to roll 2 or 12, so these values occur most infrequently. On the other hand, 7 can arise through six different combinations and will occur most frequently. If the frequency of values are plotted, it forms a triangular distribution with 7 at the peak and 1 and 12 at the base. If more die are used, the pattern is repeated but the distribution begins to form a smoother curve, the shape of which is described as "bell-shaped." This curve also describes the *normal distribution.* Although other distributions exist, most of the data we collect will be distributed normally or can be transformed into a normal distribution. Most of the more common statistical calculations require that data be normally distributed. *Non-parametric statistics* are used for data that don't occur with a normal distribution.

One of the most useful ways to describe normally distributed data is based on where the peak occurs, which ideally is in the center of the distribution. The central value around which the data values fall is termed the *central tendency* of the distribution. Although the most common central value we use is the *mean*, calculated by dividing the sum of all data points by the number of data points (Equation 1).

$$\bar{x} = \frac{\Sigma x}{n} \tag{1}$$

where: $\bar{x}$ = the mean
$\sum$ = the sum of
x = numerical value of data points
n = number of data points

One shortcoming of the mean is that it is sensitive to extreme values. A single outlying data point can skew the distribution so that the mean no longer represents the peak of the curve. Because of this, other descriptors of the central tendency may actually be more useful. The *median* is the middle point of the distribution, with an equal number of points above and below it. The median is most often used with ordinal data. The *mode* is the data value that occurs most frequently in a data set. It is the only measure of central tendency possible with data in a nominal scale. In a perfect normal distribution, all three descriptors are the same. The choice of which to use should depend on the actual shape of the distribution. Sometimes the curve will be skewed and will require mathematical transformation to better approximate the normal distribution.

Variability

Although most data points will be clustered around a central value, some will be widely dispersed. The wider the dispersion, the more difficult it is to have confidence in the data interpretations and comparisons. A quantitative measure of the spread is critical to test hypotheses. *Variance* is the average of the squared deviations, calculated as the sum of squared deviation from the mean divided by the sample size minus one (degrees of freedom) (Equation 2).

$$s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1} \tag{2}$$

where: s^2 = variance of the data
$\sum$ = the sum of
x = the numerical value of a data point
$\bar{x}$ = the mean of the data
n = the number of data points

The variance does not provide a very useful description of variation within the data set. The positive square root of the variance is known as the *standard deviation* (Equation 3), which

quantitatively describes the data spread.

$$s = \sqrt{s^2} \qquad \textbf{(3)}$$

where: s = the standard deviation
s^2 = the variance

The standard deviation represents the average amount by which each observation in a data set differs from the mean. A small standard deviation indicates that all values in the data set are very close to the mean. One standard deviation on each side of the central point is equal to 68.26% of the area under the normal curve. Two standard deviations are equal to 95.44%, three are 99.74%.

If means are calculated from several sample data sets taken from the same population, there will be distribution around the true population mean. The distribution of these means is described by the *standard error* (Equation 4).

$$SE = \frac{s}{\sqrt{n}}$$

where: SE = the standard error of the mean
s = the standard deviation
n = the number of data points

The standard error describes the variability of the means around the true population mean.

The *central limit theorem* states that sample means drawn from an arbitrarily distributed population tends to be normally distributed.

Graphs

Graphs are pictorial representations that summarize numerical data. There are many different formats but in all cases a well-designed graph will immediately display patterns in the data. Graphs should be simple, self-explanatory, and with clearly labeled variables including units of measure. *Independent variables* are plotted along the X-axis; *dependent variables* are plotted along the Y-axis.

Histograms

Histograms display a frequency distribution for discrete or continuous data. The X-axis displays intervals and the Y-axis displays the frequency of observations within each interval. For example, exam grades may be summarized with possible grades as the independent variable and the number of students earning each grade as the dependent variable. *Bar charts* are similar to histograms but are used with nominal or ordinal data.

Frequency Polygons

Frequency polygons are very similar to histograms but use a continuous line instead of bars to

represent the data. The major advantage of a frequency polygon over a histogram is the ability to view and compare several data sets on the same graph.

Box Plots

Box plots require only a single axis and may be presented in a horizontal or vertical manner. A box plot appears to be a bisected box with perpendicular lines extending from two opposite sides of the box. The lower (or left) side of the box represents the 25th percentile (or quartile). The higher (or right) side of the box represents the 75th percentile. The line that bisects the box represents the median or 50th percentile. The lines projecting out from the box extend to the extreme observations and indicate the range of observations.

Scatter Plots and Line Graphs

Scatter plots and *line graphs* display the relationship between two continuous variables. The primary different between the two is that intervals on the independent axis are scaled to actual values in the scatter plot while the intervals in line graphs may not all have the same numerical value.

Hypothesis Testing

The basis of hypothesis testing is the comparison of observed results with expected results. Statistics allow us to determine if an observed result is significantly different from an expected result. The most common question is whether the means two groups are significantly different. In statistics. *significant* is a very specific term, used to denote statistical determinations. Stating that two means are significantly different means that the difference is greater than could be attributed to the natural variability of the population or the method.

In order to determine if a hypothesis should be accepted or rejected, it is necessary to select a level of significance (α). The most common level in biological sciences is $\alpha = 0.05$. At that level, 5% of the time a true null hypothesis will be rejected. Another way of considering this is that the investigator can be 95% confident that a true null hypothesis will not be rejected. Rejecting a correct null hypothesis is called a *type I error*. Accepting a false null hypothesis is called a *type II error.* Decreasing the α lowers the probability of a type I error but raises the probability of a type II error.

The selection of the appropriate statistical test and proper experimental design requires several steps. The investigator must be clear on the problem being addressed and the type of data (measurement scale and distribution) expected. This will allow selection of the appropriate statistic, which in turn allows the hypotheses to be stated. Common statistical procedures are summarized in Figure 6.1. Once it is known how the data will be used, the experiment can be designed and performed. Analysis of the data will determine if the null hypothesis or the alternative hypothesis will be accepted.

Parametric Tests

Parametric tests of significance are used with data that are normally distributed. Data that cannot be assumed to follow a normal distribution need to be analyzed with non-parametric tests. Parametric tests are more powerful, that is they are less likely to result in a type II error.

The Student t Test

The Student t test examines whether the means of two groups are the same. The *t* distribution varies as a function of degrees of freedom and is symmetrical about a mean of 0. The more degrees of freedom, the less spread out the distribution. To use the *t* test, a *t* value calculated from the data is compared to a critical value based on the *t* distribution with degrees of freedom determined from the sample sizes. The *t* value is calculated by dividing the difference of the two means by the standard error of the difference of the means. The formula used depends on whether the population means are identical, which is the usual assumption (Equation 5).

where:		
t	=	t value
$\bar{x}_A$	=	mean of group A
$\bar{x}_B$	=	mean of group B
$\sum(x_a - \bar{x}_A)^2$	=	sum of squares for group A
$\sum(x_a - \bar{x}_B)^2$	=	sum of squares for group B
n_A	=	number of data points in group A
n_B	=	number of data points in group B

The null hypothesis is accepted if the calculated *t* value is equal to or less than the tabulated *t* value.

The Student's *t* test can only be used to compare two means. A variation uses paired data, but this increases the chance for a type I error. The analysis of variance procedure with multiple range tests may be a better choice for use with multiple groups.

Analysis of Variance

Hypothesis testing by comparing multiple groups is best done with an analysis of variance (ANOVA). The *F* statistic is calculated and compared with tabulated values from an *F* distribution. Usually a control group is compared to several experimental groups. The results only indicate if significant differences exist and not the magnitude or direction of differences.

Two independent variances are calculated, between-group variance and within-group variance. The *F* statistic is the ratio of between-group divided by within-group variance (Equation 6).

$$F = \frac{\Sigma n_i((\overline{x_i} - \overline{x})^2 / (k - 1))}{\Sigma\Sigma(x_{ij} - x_i)^2 / (n - k)} \qquad \textbf{(6)}$$

$$t = \frac{(\overline{x_A} - \overline{x_B})}{\sqrt{\left(\frac{\Sigma(x_a - \overline{x_A})^2 + \Sigma(x_b - \overline{x_B})^2}{(n_A + n_B - 2)}\right)\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}} \qquad \textbf{(5)}$$

where:		
F	=	the *F* ratio
x_{ij}	=	the *j*th observation in Group *i*

$\bar{x}_i$	=	the mean of Group *i*
$\bar{x}$	=	the overall mean
k	=	the total number of groups
n	=	the number of observations
n_j	=	the number of observations in Group *i*

Each variance has its own degrees of freedom. The degrees of freedom of between groups is the number of groups minus one. The number of degrees of freedom of the within groups is the total sample size minus the number of groups. If the *F* ratio is greater than the critical value from the table, the null hypothesis is rejected.

Nonparametric Tests

If the data cannot be assumed to be normally distributed and cannot be transformed into a normal distribution, nonparametric tests must be used. These tests are not as powerful as parametric tests but are more conservative and require less assumptions. Some of the more common tests are listed in Figure 6.1. The rank sum or Mann-Whitney test is analogous to the *t* test. The nonparametric alternatives to the paired *t* test are the sign test, for when the magnitude of differences cannot be determined, or the Wilcoxon matched-pairs test when the differences can be ranked by the magnitude of the difference. The Kruskal-Wallis test replaces ANOVA with independent samples. Related samples are examined with the Friedman test.

Correlations and Regressions

This statistic calculates the degree of relatedness between two parameters of a population. When plotted, a positive correlation displays a positive slope. A negative correlation (as one parameter increases the other decreases proportionally) displays a negative slope. If the two parameters are unrelated, the plot is a collection of random points without an obvious pattern. The strength of the relationship between two variables is reported as the correlation coefficient, symbolized as *r*. Correlation coefficients can be calculated in a number of ways but most calculators and spreadsheets use Pearson's method. The value can vary between -1 and +1, corresponding to the slope displayed by the relationship.

Regression analysis defines a mathematical relationship between two parameters that permits prediction of one parameter if the other is known. Once data is plotted, the investigator tries to determine the type of relationship that the data fit. A regression line is constructed so as to minimize the variance of the data about the line. The most common form is linear regression, where a straight line is described by the equation $y = mx + b$, where m is the slope and b is the y-intercept. Nonlinear regressions are more difficult to calculate and may require a fairly powerful curve fitting program. The equation for the regression line allows the investigator to calculate corresponding values given *x* or *y*.

Regressions have two sources of variance. In addition to the variance of the dependent variable, there is also an error variance that arises from the relationship between the independent and dependent variable. Without an error variance, all observations would fall on the regression line. The *coefficient of determination*, symbolized by r^2, is the measurement of the variance due to the relationship of the dependent variable to the independent variable. A value of 1 indicates that the value of the dependent variable completely depends on the independent value.

6 USING A DISK DIFFUSION ASSAY TO INTRODUCE STATISTICAL METHODS

Background

Antibiotic resistance is becoming an ever-increasing concern. Thus, the ability to accurately assess the level of resistance of clinical isolates is of great importance, ultimately affecting patient outcome. Antibiotic resistance testing using the disc diffusion technique was developed in the 1940s soon after the discovery of the first antibiotics. In 1966, Bauer *et al.* (1) published their paper which helped standardize this protocol. Today, NCCLS provides periodic updates of this protocol (6) and tables (5) so when performing this assay results are reproducible from day to day and from lab to lab if using the same isolate and the same antibiotics. That is, if the directions are followed, the same results will always be obtained with the same isolate, regardless of where or when the analysis is performed.

In this exercise we modify the NCCLS disc diffusion assay so that various concentrations of the inoculum are used in addition to the standard inoculum concentration. After incubation for each inoculum, the diameter of the zone of inhibition of each antibiotic disc is measured. Comparisons are made of the effect of inocula concentration on the diameter of the zone of inhibition. By testing this variable (i.e. different concentrations of inocula) we hope to show the importance of following a standardized protocol and to illustrate the use and the value of statistical analysis in microbiology. The null hypothesis, that is the hypothesis of no difference, is that the diameters of the zones of inhibition will be the same regardless of *E. coli* concentration in the inocula. The hypothesis is tested using the *t*-test. In addition, each student in the group measures the zones of inhibitions and their results are compared. In this instance, the null hypothesis, that there is no difference between their measurements, is tested with ANOVA.

Students should read the protocol and organize their groups before the start of the experiment. In preparation for this exercise, we recommend that the instructor have them address the following points:

- Draw a flow diagram of the experiment noting how you would make the dilutions and then perform the plate counts.
- This experiment investigates the effect of bacterial concentration in the disk diffusion assay. What other variables could be investigated and which statistical tools would be appropriate for analysis?
- What are the modes of action of the antibiotics used? Would they be expected to act against gram positive, gram negative organisms or both?
- Do you think that it would make a difference if more than one person inoculated the plates for antibiotic testing? What statistical test would you use to test your hypothesis?
- How is resistance or sensitivity to an antibiotic determined using the disc

diffusion assay? Does this have clinical relevance? What is the measurement that is used to determine this? Would the a measurement of the same magnitude imply resistance (or sensitivity) if the organism tested was a different species?

- Would you expect bacteria in the stationary phase of growth to be more sensitive or resistant to an antibiotic as compared to the same species in the exponential phase? (Hint: Would pre-incubation of plates with the antibiotic disks in place affect the size of the zone around the disk?)
- List some ways in which antibiotic resistance can be transferred from one bacterium to another. Briefly describe each mechanism.

Purpose

At the completion of this activity, students will

1. understand the importance of using standardized methods, *i.e.* antibiotic resistance analysis using the disc diffusion method, and recognize important variables involved in antibiotic resistance analysis
2. better understand the scientific method and its application to experimental investigations
3. be more familiar with hypothesis testing and the use of inferential statistics
4. be able to use spreadsheets for simple statistical functions
5. have improved various lab skills, such as aseptic technique, preparing dilutions, making viable counts, inoculating spread plates, and using pipettes

Materials (per 4 students)

32 Mueller Hinton agar plates
sterile cotton swabs
one ml sterile pipettes
sterile tubes
test tube racks
overnight culture of *E. coli* on TSA
E. coli culture adjusted to 0.5 McFarland Standard
tubes of sterile saline for dilutions (4 x 9.9 ml, 6 x 9.0 ml, plus some for adjusting culture to McFarland standards)
McFarland standards (0.5, 1, 2,4)
gentamicin antibiotic disks (GM10)
sulfamethoxazole/trimethoprim antibiotic disks (SXT)
forceps
95% ethanol in a beaker
Bunsen burner
vortex
ruler or caliper
magnifier

sterile bent plastic rods... “hockey sticks”
ice in small buckets or beakers

Procedure

Day 1 (Inoculations)

1. **Divide into groups of 4 students each.**
2. Each group will be given a culture diluted to 0.5 McFarland standard. This is the cell concentration recommended in the NCCLS protocol and will be considered the control culture in this experiment. You will also be assigned another cell concentration based on the McFarland standard (1, 2, or 4) and will need to make the appropriate suspension from an *E. coli* plate culture using sterile saline. Aseptically transfer colonies from the plate into a tube of sterile saline, vortexing after each addition until the colony is completely dispersed. Continue adding cells until the turbidity of your suspension matches the appropriate McFarland standard. (Hint: Looking at a line of print through the standard and the cell suspension is a good way to compare the density.) The cell suspension you make will be considered the experimental culture. Keep both cultures on ice to inhibit growth and keep cell concentrations from changing during the course of the experiment.
3. Each group will need two tubes of 9.9 ml sterile saline and three tubes of 9 ml sterile saline to enumerate each of the two cell cultures. (Four tubes of 9.9 ml and six tubes of 9 ml saline, total.) Label each series of tubes with the culture density (0.5 McFarland and your other density). Label the first 9.9 ml tube 10^{-2} and the second 9.9 ml tube 10^{-4}. Label the three 9 ml tubes 10^{-5}, 10^{-6}, and 10^{-7}. These tubes will be used to dilute the control and experimental culture to make inocula for spread plates.
4. Plate counts should be performed by diluting the two cultures with the saline dilution blanks (step 3, above). Start by transferring 0.1 ml from the control culture to the tube marked “10^{-2} ” using good aseptic technique, then vortex that dilution. Using a new, sterile pipette, transfer 0.1 ml from the 10^{-2} dilution to the tube marked “10^{-4}” and vortex that tube. Transfer 1 ml from the “10^{-4}” tube to the tube labeled “10^{-5}.” Continue in this manner, inoculating the last two dilutions. Repeat this entire process with the experimental culture.
5. Use your dilutions to inoculate spread plates in order to determine the viable cell count in your cultures. For the control culture and experimental cultures adjusted to the 1 McFarland standard, inoculate duplicate plates each with 0.1 ml from the 10^{-4}, 10^{-5}, and 10^{-6} dilutions. With the experimental cultures adjusted to the 2 or 4 McFarland standard , inoculate duplicate plates each with 0.1 ml from the 10^{-5}, 10^{-6}, and 10^{-7} dilutions. Use a sterile hockey stick to spread the inoculum over the entire surface of the plate, using a fresh hockey stick for each dilution with each culture.
6. For the disk diffusion assay, each group should swab ten Mueller-Hinton agar plates from the control culture. Repeat this, inoculating ten more Mueller-Hinton

agar plates from the experimental culture. Swabbing should be done by following the NCCLS protocol (2, 6) :

A. Dip a sterile cotton swab into the appropriate liquid culture.
B. Press the swab against the inside of the tube to squeeze out excess fluid.
C. Lightly rub the swab over the entire surface of the Mueller-Hinton plate.
D. Rotate the plate 60º and swab the surface a second time.
E. Rotate the plate 60º and swab the surface a third and final time.
F. Roll the swab on the agar around the inside edge of the plate.

You can use the same swab to inoculate every plate in the series but you must dip it in the liquid culture before swabbing each plate.

7. After five minutes, but no more than 15 minutes, the two different antibiotic discs should be placed on the agar surface of each of the ten plates from each culture, at least 24 mm from each other and 10 mm from the edge of the plates, and tapped down with sterile forceps. (Placement may be done using an automatic dispenser or by using loose disks in a sterile petri plate, placing the disks with sterile forceps.) Sterilize forceps by dipping the tips in 95% ethanol and igniting the ethanol by rapid passage through a flame. (Keep the tip pointed down, away from the ethanol reservoir, and do not hold the forceps in the flame!) Invert the plates and incubate at 37º C for 18 hours.

Day 2 (Record Data)

1. After 18 hours, the plates with the antibiotic disks should be removed from the incubator and the diameters of the zone of inhibition should be measured in mm by each student in the group (*i.e.,* four sets of measurements for the control culture and four sets of measurements for the experimental culture). The use of a magnifier, such as those on colony counters, greatly improves accuracy. Record the results in your lab notebook and share your data with the other members of your group.
2. Count the colonies on the enumeration plates, recording your counts. Calculate the viable cell concentration in the two McFarland cultures and determine if their ratio matches the ratio between the McFarland values for the cultures.

Day 3 (Analyze Data)

1. Determine if the sizes of the zones of inhibition with the two McFarland cultures are the same or different for each antibiotic using the t-test. Typical student data is shown in Table 1. This can be done manually using a calculator or with a spreadsheet (recommended, see Appendix). The null hypothesis (no difference between the diameters of the zones of inhibition for different concentrations of *E. coli* in the inocula) is accepted if the calculated *t*-value is less than the critical *t*-value with an alpha of 0.5 and 18 degrees of freedom. A one-tailed test is

appropriate since a higher concentration of *E. coli* may produce a zone of inhibition that is equal to or smaller than a zone resulting with a lower concentration of cells. Thus the critical value of *t* is 1.734.

2. Determine if the individual measurements of the zones of inhibition made by each group member are the same or different for each antibiotic using ANOVA. Student data for a control culture with trimethoprim/sulfamethoxazole is shown in Table 2. The null hypothesis that there is no significant difference between measurements of the diameter of the zones of inhibition made by different students is accepted if the calculated *F*-value is less than the critical *F*-value with an alpha value of 0.5 and 3 degrees of freedom, that is 2.866.

Results

In addition to recording data in lab notebooks, use a spreadsheet for analyses and use those data sheets as the basis of a report. The report should include the raw data, results of statistical analyses, and interpretations of the statistical analyses. Describe, in your own words, the importance of a standardized antibiotic testing protocol.

A tutorial on using Microsoft® Excel for this experiment is available online on the instructor's site (http://faculty.weber.edu/wlorowitz/micro_procedures.htm). Samples of data are shown in Tables 1 and 2. Table 1 displays typical student data for diameters of zones of inhibition from gentamicin for *E. coli* cultures adjusted to 0.5 and 1 McFarland standards. Results of the *t*-test, performed using Microsoft® Excel, indicate that the zones with the greater cell density were significantly smaller than the zones with the less dense culture. Table 2

Table 1. Comparison of zones of inhibition produced from gentamicin with *E. coli* cultures adjusted to 0.5 and 1 McFarland standard turbidities. The *t*-test was used to examine the effect of cell density on antibiotic inhibition.

Disk Diffusion Assay						
GM10, 0.5	GM10, 1			t-Test: Two-Sample Assuming Equal Variances		
24	24					
25	26				*GM10, 0.5*	*GM10, 1*
24	25			Mean	24.4	25.1
25	24			Variance	0.266667	0.544444
25	25			Observations	10	10
24	26			Pooled Variance	0.405556	
24	25			Hypothesized Mean Difference	0	
25	26			df	18	
24	25			t Stat	-2.457864	
24	25			P(T<=t) one-tail	0.012172	
				t Critical one-tail	1.734063	
24.4	25.1	= mean		P(T<=t) two-tail	0.024345	
0.516398	0.737865	= s.d.		t Critical two-tail	2.100924	

displays typical student data for measurements of the diameters of zones of inhibition from gentamicin for an *E. coli* culture adjusted to a 0.5 McFarland standard, made separately by four different students. ANOVA (using a spreadsheet) suggests no significant difference between the sets of measurements.

Table 2. Measurements of the diameter of zones of inhibition made with the same ten samples by each of the four group members. Data were compared using ANOVA.

Disk Diffusion				
SXT, 0.5				
A	**B**	**C**	**D**	**= student**
29	29	30	27	
32	29	31	31	
29	29	30	27	
30	32	31	31	
30	31	31	30	
30	32	30	31	
30	31	31	30	
30	30	30	30	
29	29	30	31	
29	30	30	30	
29.8	30.2	30.4	29.8	= mean
0.918937	1.229273	0.516398	1.549193	= s.d.

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
A	10	298	29.8	0.844444
B	10	302	30.2	1.511111
C	10	304	30.4	0.266667
D	10	298	29.8	2.4

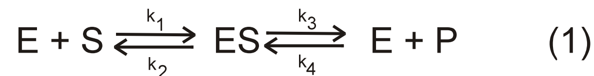
ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	2.7	3	0.9	0.716814	0.548441	2.866265
Within Groups	45.2	36	1.255556			
Total	47.9	39				

7 MOCK ENZYME KINETICS

Background

Enzyme kinetics describes the rate of change of reactant concentrations in a chemical reaction. A simple, one-substrate reaction can be described by Equation 1, where E is the enzyme, S is the substrate, and P is the product:



$$E + S \underset{k_2}{\overset{k_1}{\rightleftharpoons}} ES \underset{k_4}{\overset{k_3}{\rightleftharpoons}} E + P \qquad (1)$$

Initially there is no product, so the rate constant k_4 is zero. Under steady-state conditions, the concentration of ES stays the same while the concentrations of the reactants and products are changing; that is, the rate of ES formation is the same as the rate of breakdown (Equation 2).

$$k_1[E][S] = (k_2 + k_3)[ES] \qquad (2)$$

Equation 2 can be rearranged and simplified by defining the Michaelis constant, *K*m (Equation 3):

$$K_m = \frac{k_2 + k_3}{k_1} \qquad (3)$$

When all of the enzyme is bound to substrate (*i.e.,* substrate concentration is at saturation levels) and is thus in the ES state, the rate of the reaction is at a maximum or V_{max}. Combining K_m with V_{max}, Michaelis and Menton derived a mathematical equation (Equation 4) to describe the relationship between the initial velocity (or reaction rate) and substrate concentration:

$$v_o = \frac{V_{max}\,[S]}{K_m + [S]} \qquad (4)$$

Figure 1 illustrates a typical plot for a one-substrate enzyme-catalyzed reaction. At relatively low substrate concentration, v_o increases as the substrate concentration increases. As v_o approaches V_{max} the reaction rate is constant and independent of substrate concentration.

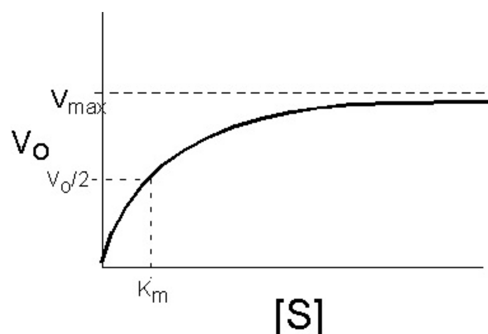


Figure 1. Michaelis-Menton plot showing the relationship of substrate concentration and initial velocity in enzyme-catalyzed reactions.

Rearranging the Michaelis-Menton equation to the form in Equation 5 makes it apparent that the Michaelis constant, K_m, has the same units as substrate concentration or moles/liter. It should also be apparent that when K_m is the substrate concentration, the initial velocity, v_o, is equal to one-half the maximal velocity, V_{max}. (That is, using Eqn. 5, when $K_m = [S]$, $V_{max}/v_o = 2$.) K_m can give an approximation of the affinity (think of affinity as the strength of attraction between substrate and enzyme) that the enzyme has for the substrate; the lower the value of K_m, the greater the affinity.

$$K_m = [S] \left(\frac{V_{max}}{v_o} - 1\right) \quad (5)$$

Although the values of K_m and V_{max} can be estimated from plots like that in Figure 1, it is easier and more accurate (without using nonlinear regression) to use a plot of a linearized form of the Michaelis-Menton equation. The reciprocal of the Michaelis-Menton equation (Equation 6) is the most common linearization and the concomitant plot (Figure 2) is referred to as a double-reciprocal or Lineweaver-Burk plot.

$$\frac{1}{v_o} = \frac{K_m}{V_{max}} \frac{1}{[S]} + \frac{1}{V_{max}} \quad (6)$$

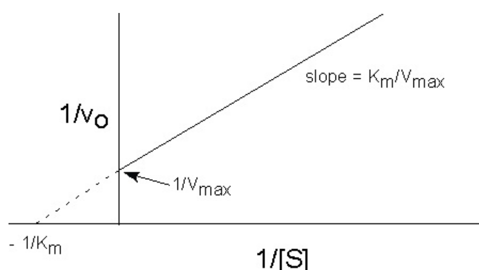


Figure 2. Lineweaver-Burk (double-reciprocal) plot.

By comparing Equation 6 to the general equation for a straight line (y = a+ bx), it is clear that the slope of the line is K_m/V_{max} and the y-intercept is $1/V_{max}$. The x-intercept is equal to $-1/K_m$. It is important to note that a shortcoming of using a double-reciprocal plot is that at low substrate concentrations (high values for 1/[S]) small deviations in the initial velocity (v_o) can badly skew the regression line through the data points. For this reason other linearized forms of the Michaelis-Menton equation, for example Equation 7, used to produce an Eadie-Hofstee plot (Figure 3), may be used. The Eadie-Hofstee plot is not without its own problems, though.

$$v_o = -K_m \frac{v_o}{[S]} + V_{max} \qquad (7)$$

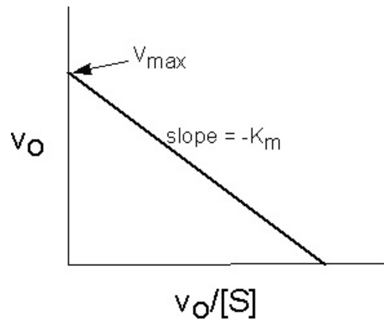


Figure 3. Eadie-Hoffstee plot.

As Equation 7 indicates, the "independent variable" (*x*-axis) actually comprises the true independent variable ([*S*]), as well as the dependent variable (v_o).
The values for K_m and V_{max} can be determined using either linearized form**.** The calculated values will not be identical but they should not be very different.

Competitive Enzyme Inhibition

In competitive inhibition, an inhibitor molecule has a chemical structure similar to the substrate and competes with the substrate to attach to the catalytic site of the enzyme. The effectiveness of a competitive inhibitor is based on its relative concentration to the substrate and the affinity with which it binds to the enzyme. Inhibition increases as the ratio of inhibitor concentration to substrate concentration increases and/or as affinity of the enzyme for the inhibitor increases (effective binding at low concentrations). An inhibitor constant, K_I, similar to the Michaelis constant, can be defined (Equation 8) and incorporated into the Michaelis-Menton equation (Equation 9).

$$K_i = \frac{[E][I]}{[EI]} \qquad (8)$$

$$v_o = \frac{V_{\max}\,[S]}{K_m\left(1 + \frac{[I]}{K_I}\right) + [S]} \qquad (9)$$

As before, an easy way to determine the value of K_I is by taking the reciprocal of Eqn. 9 (Equation 10) and making a double reciprocal plot (Figure 4). As with the uninhibited enzyme,

$$\frac{1}{v_o} = \frac{K_m}{V_{\max}}\left(1 + \frac{[I]}{K_I}\right)\frac{1}{[S]} + \frac{1}{V_{\max}} \qquad (10)$$

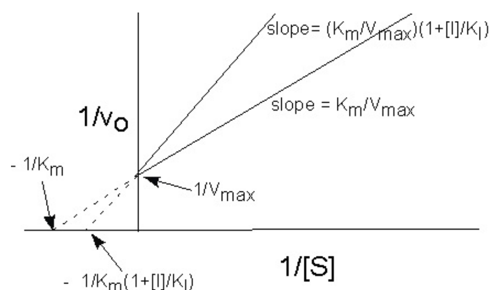


Figure 4. Lineweaver-Burk plot with competitive inhibition.

the y-intercept represents $1/V_{\max}$. However, the apparent K_m is greater due to the inhibition, so the slope is steeper and the x-intercept closer to the origin. By plugging in the values of K_m and V_{max} from the uninhibited enzyme, the value for K_i can be determined using Equation 10 and values from the plotted data. Alternatively, K_i can be calculated from the slope of the line generated from the inhibited enzyme.

Mock Enzyme Lab

(after Chayoth, R. and A. Cohen. 1996. A simulation game for the study of enzyme kinetics & inhibition. American Biology Teacher 58:175-177)

PROCEDURE

Part I Enzyme Activity and Substrate Concentration

A blindfolded student (the "enzyme") picks up as many pennies ("substrate") as possible in 30 seconds. At lower substrate concentrations (*e.g.,* 10 pennies) you may want to shorten each reaction time to 10 seconds. (Why?) Alternatively, you may want to replace pennies, keeping their concentration constant. Make sure to record the method you choose and include that information in your lab report.

Different numbers of pennies (10, 20, 30, . . ., 100) are placed randomly within a 1 m x 1 m square. (What are your units for substrate concentration?) The blindfolded student picks up as many pennies as possible within the allotted time, placing them within a cup in her other hand. Normalize the data by calculating the rate in minutes (*e.g.,* for a 30 second assay, multiply by 2).

Plot the data as it is collected as a Michaelis-Menton plot and as at least one linearized form. This is termed "real-time data analysis" and ensures that you will have good data for analysis. (No excuses for bad data! Do it right the first time and you won't have to repeat this experiment on your own time.)

Part II: Competitive Inhibition

The simulation in Part I is repeated, with the addition of silver pennies, representing a competitive inhibitor. The concentration of the inhibitor is kept constant while substrate concentrations are varied. At the end of each assay, only the copper-colored pennies are counted.

Repeat Part I, including 10 silver pennies with each trial. Repeat using 30 silver pennies.

Tips:

1. The "enzyme" should not get better (or worse) with practice.
   - try to use a steady rhythm
   - use your catalytic site (finger tips) to find the pennies, not your palm
   - aim for consistency

- Get enough data points for a smooth Michaelis-Menton curve and to give a good distribution on your linearized plot.
  - pay special attention to the transition phase (mixed order region) of the

Michaelis-Menton plot
- consider replicates
- be especially diligent at low substrate concentrations with the Lineweaver-Burk plot

3. Randomize your substrate concentrations (don't do an orderly progression).

RESULTS

1. Create a table similar to the one below for each trial (uninhibited, two trials with competitive inhibition); substitute or include columns for an Eadie-Hofstee plot if that is your choice:

[S] (coins/m^2)	v_o (coins/30 sec)	v_o (coins/min)	1/[S]	1/v_o

2. Plot as you go.

DISCUSSION

1. Analyze your data using Microsoft® Excel. (Described in the online tutorial.)

2. Your lab report should include
   a. values for K_m, V_{max}, and K_I. (with descriptions in "Materials and Methods" section of how values were obtained)
   b. plots of each assay (easiest to do as separate plots)
   c. regression information for each plot including the equation of the line and the r^2 value
   d. discussion of your data in regard to how well your "enzyme" worked, the values obtained from your plots (including regression results), and the effectiveness of the competitive inhibitor

8 ENUMERATION

Determining the number of cells is important in a wide range of applications, *e.g.,* determining optimum growth parameters, measuring efficiency of antimicrobials, or environmental monitoring. Counts can be *direct*, where the cells themselves are counted, or *indirect*, where a property of the cells is measured. In addition, enumeration can involve all cells in a *total count* or only the live cells in a *viable count*.

Direct Counts

Total Counts

Microorganisms can be enumerated by using a microscope to actually count the number of bacteria contained within a standard, known volume. In general, a suspension of microorganisms is added to a specially designed and precisely manufactured microscope slide that is etched with a grid of known dimensions and uses a rigid coverslip to provide a uniform volume. Several types of **counting chambers** exist; the **Petroff-Hausser counting chamber** is the most common one used in bacteriology while **hemocytometers** are used to make blood cell counts. By counting the cells within the grid (a known area) and multiplying that by the distance between the slide and coverslip, it is possible to calculate the concentration of cells per known volume. Prior to counting cells, it is almost always necessary to dilute the original sample.

Direct microscopic counts can also be made by staining cells with a dye and putting the sample through a membrane filter. Fluorescent dyes make this a very useful technique with environmental or other types of "dirty" samples.

The most common automated method for making total counts is with a Coulter counter. This instrument sips a set volume of a cell suspension and passes it through a narrow chamber with two electrodes that are used to measure the electrical resistance of the fluid flowing past. Resistance increases when a cell passes between the electrodes. Enumeration with a Coulter counter is fast and accurate as long as the medium does not contain non-cellular particles.

Viable Counts

Viable counts are especially useful when trying to determine the effectiveness of an antimicrobial, for example a disinfectant. Although fluorescent dyes are available that differentially stain metabolizing (live) and non-metabolizing (probably dead) cells, the most common way to obtain viable counts is by counting colonies on spread or pour plates. Plates used for viable counts must have between 30 and 300 colonies (150 mm diameter plate) for proper determinations. The lower limit is set on the basis of statistical accuracy and the upper limit is set on the basis of competition and overcrowding. As with microscopic counts, samples must be diluted.

Indirect Counts

Total Counts

Total cell numbers can be estimated from a property of cells, for example the concentration of protein, the dry weight, or the amount of light absorbed. In all cases, a *standard curve* must be

created to accurately associate the measured parameter with cell number.

Although cell features like dry weight or protein concentration can provide very accurate estimates of cell concentrations, the methods to obtain data are somewhat tedious and, like viable counts, may require quite a long time before the data is available. However, these methods can provide additional data, for example dry weights can be used to help determine growth yields.

The most common way to monitor growth is by measuring the amount of light absorbed by a culture. A culture tube is inserted in a spectrophotometer set at a specific wavelength (usually around 600 nm) and the absorbance is measured. As the cells grow, the culture becomes proportionally more turbid, providing an excellent relative measure of growth. This method is simple to perform, provides data within seconds, and is noninvasive.

Viable Counts

The most common method to indirectly measure viable cells is with the Most Probable Number (MPN) technique. The MPN technique estimates the number of viable cells in a sample using mathematical inference based on the number of tubes inoculated with different sample dilutions that show growth. Typically, MPN measurements are made with sets of three, five, or ten replicates at each dilution due to the availability of prepared interpretation tables.

Dilutions

In many situations the concentration of microorganisms are too high to perform direct counts. When this occurs, samples are diluted to workable levels. Dilutions are usually made in factors of 2 or 10, usually through a series. For most lab applications, diluents are dispensed in tubes so that the final volume (diluent + sample) is 10 ml. For example, 5 ml of *diluent* would be used for 1:2 dilutions, whereas 9 ml diluent and 9.9 ml diluent would be used for 1:10 and 1:100 dilutions, respectively. There isn't anything magic about using 10 ml volumes, they're just a convenient size to store and use. Besides using good aseptic technique, it's important to use precise volumes with dilutions since even a small aberration can be multiplied several-fold during a serial dilution.

The mathematics of dilutions is straightforward but can be confusing if you just try to "fudge" your way through. Remember to keep track of your units (dimensional analysis) and learn the relationships involved in making dilutions. First, consider how *concentration* is defined:

Concentration = Particles/Volume

or, $$\mathbf{C = P/V} \tag{7.1}$$

For example, if 0.1 ml of culture is used for a spread plate and 250 colonies are counted:

C = 250 CFU/0.1 ml =2500 CFU/ml

Where: C = the concentration of colony forming units (CFU) per ml in the culture plated
P = the number of CFU on the plate
V = the volume plated

However, what if the aliquot plated was from a 1/100 dilution of the culture? Of course, the concentration of CFU in the original culture would be 100X higher than the value calculated above, or 150,000 CFU/ml. This is an obvious relationship and seems somewhat intuitive but we can define the relationship mathematically.

When a suspension is diluted, the number of particles doesn't change:

$$\mathbf{P_1 = P_2} \qquad (7.2)$$

but, since the volume increases, the concentration (P/V) decreases (the whole purpose of performing dilutions). By rearranging the equation for concentration (Eqn. 7.1) above:

$$\mathbf{P = CV} \qquad \mathbf{(7.3)}$$

Since the number of particles is the same in each sample, the concentration in each sample (expressed, for example, as CFU/ml) times the volume (ml) would be the same.

$$\mathbf{C_1V_1 = C_2V_2} \qquad (7.4)$$

To calculate dilutions:

$$Dilution = \frac{Volume\ added}{Total\ volume} = \frac{Volume\ added}{Volume\ of\ diluent + Volume\ added} \qquad (7.5)$$

For example, if a culture contains 1×10^9 cells/ml and 0.1 ml is added to 9.9 ml of diluent,

***Dilution* = 0.1ml/(9.9 ml + 0.1 ml) = 0.1 ml/10 ml = 0.01 or 1/100**

The *Dilution Factor (DF)* is the reciprocal of the dilution, or **1/dilution**.
The *Total Dilution* is the product of all dilutions:

$$\boldsymbol{D_{total} = D_1 \times D_2 \times D_3 \ldots}$$

What does all this mean for spread plates and pour plates? The initial concentration, $\mathbf{C_1}$ (that is, the concentration of cells in the undiluted culture) is equal to $\mathbf{C_2}$ (the number of colonies/volume plated, or $\mathbf{P_2/V_2}$) x **DF**, or:

$$C_1 = C_2DF = (\frac{P_2}{V_2})DF \qquad (7.6)$$

where C_1 = initial concentration (or concentration in the stock culture)
C_2 = CFU/ml (or the concentration of CFU on the plate)
P_2 = CFU on the plate
V_2= Volume plated (ml)
DF = Dilution Factor

Optical Density Measurements

Background

Microbial concentrations can be measured indirectly in a number of ways, the most common relying on light scattering by a suspension of cells. Somewhat simply defined, light scattering is what occurs when light energy strikes an object and is not absorbed. The light must be re-radiated and can emerge in any direction. The light scattered from a small particle depends inversely on the fourth power of the wavelength of light. Therefore, small particles scatter blue light more strongly and transmit red light more efficiently. (That's why the sky is blue except when the sun is below the horizon.) Microbes are large enough so that light scattering most nearly approximates an inverse second power, and therefore appear whitish (less separation of color).

The most common instrument from measuring optical density is a *spectrophotometer*. Essentially, light of a specific wavelength is directed through a microbial suspension and a photocell measures the amount of light that comes through. The intensity of the light striking the photocell is measured by a galvanometer connected to a display (analog or digital), with a result we interpret as *absorbance* (A) or *optical density* (OD). Most colorimeter and spectrophotometers also have a scale for the amount of light transmitted (%T). Although this can be a useful unit at high turbidities (converting the values to absorbance), microbiologists tend to use absorbance since this is directly proportional to the cell density of a suspension. In colorimeters the wavelength of light is set using filters, whereas spectrophotometers use prisms or defraction gratings.

Absorbance is defined as the logarithm of the quantity expressed by dividing the incident light by the transmitted light. That is,

$$A = \log (I_o/I)$$

where: I_o = the incident light
I = the transmitted light.

This relationship comes from *Lambert's law*, which states that the proportion of incident light absorbed by a medium is independent of its intensity and that each successive unit layer of the medium absorbs an equal fraction of the light passing through it. Thus,

$$A = \log (I_o/I) = C_o/C$$

where: C_o = concentration of light absorbing material in a standard
C = concentration of light absorbing material in a sample

What this means to us, is that **the relationship between absorbance and cell concentration is linear!** That is, **if you double the number of cells, you double the absorbance. An important point to remember** is that in reality, this relationship is **not true** at high concentrations of cells and that the absorbance becomes increasingly **less** than what the formula predicts. This is due to more light getting through than expected because light scattered from one bacterium is rescattered by another bacterium so that the light is redirected back into the phototube. Furthermore, there is intercellular interference with Brownian motion, resulting in a more evenly distributed cell suspension which scatters less light away from the beam.

As stated above, turbidometry is an indirect method for microbial enumeration. Indirect refers to the fact that it is not microbial cells that are being enumerated directly, but their ability to scatter light. In order to determine the actual number of microbes, a *standard curve* must be generated in which microbial concentrations are measured directly (as with a counting chamber) and plotted against the absorbance of the culture from which the sample was counted. It follows, therefore, that absorbance is a measure of total cell concentration.

Purpose

To become familiar with the principles and use of a spectrophotometer

Procedure

Using different concentrations of dyes, read absorbance at different wavelengths of light. Note the linear relationship of concentration to absorbance and that the spectrophotometer needs to be re-zeroed for each change in wavelength.

Direct Microscopic Counts with a Hemocytometer

Background

Microorganisms can be enumerated by actually counting the number of bacteria contained within a standard, known volume. Several types of counting chambers exist; the Petroff-Hausser counting chamber is the most common one used in bacteriology. For this exercise, we will use hemocytometers, which are similar to, but contain a larger volume than Petroff-Hausser chambers. This method of quantitation produces a total cell count (as opposed to a viable cell count). A total cell count includes viable (living) and non-viable organisms. Therefore, viable cell counts are either equal to or less than total cell counts.

Counting chambers are carefully constructed slides which have a grid etched into an area contained within a small enclosure. With the addition of a flat, rigid coverslip, a sealed volume is contained above the grid. It is assumed (although it is often not the case) that the volume is exact. By counting the cells within the grid, one is able to calculate (by knowing the grid and chamber dimensions) the number of cells/ml. Hemocytometers we use are 0.1mm deep and divided into 9 squares of 1 mm^2. These squares are further divided into 25 smaller squares, which are further divided into 16 squares. Therefore, there are 400 (25 x 16) small squares/mm^2 (i.e., each small square is 1/400 mm^2).

The major advantages of this method are:

1. It's easy and relatively inexpensive.
2. It provides additional information (e.g. morphology)

Major disadvantages are:

1. It is difficult to reproducibly fill the counting chamber
2. Bacteria can adsorb to coverslips, pipettes, etc.
3. Motile cells can be difficult to count and lead to errors.

Purpose

To make direct counts of a *Saccharomyces* culture using a hemocytometer.

Materials and Methods

Saccharomyces cerevisiae
hemocytometer

1. Suspend yeast cells and make appropriate dilutions (probably 1/10 and 1/100).

2. Place coverslip on hemocytometer and use pasteur pipet to fill both chambers at introduction point.

3. Focus on central square at 100X. Count cells at 400X in at least five (**5**) secondary squares (1/25 mm^2), *i.e.,* the 4 corners and the center squares. <u>For cells on lines, count left and top</u>, not right and bottom. Buds of equal size count as 2, not 1 cell.

4. Calculate the concentration:

   (# cells in 5 squares) x 5 = # cells/mm^2 surface area **(include DF, if necessary)**

   (# cells/mm^2) x 10 = # cells/mm^3 (volume, since chamber is 0.1 mm deep)

   (# cells/mm^3) x 1000 = # cells/cm^3 or # cells/ml

Putting it all together:

 # cells/ml = (# cells in 5 squares) x (5 x 10^4) x DF

Results

Values for the total concentration of yeast cells.

Viable Counts of <u>Saccharomyces cerevisiae</u>

Background

By definition, **viable cells** are able to divide, and thus give rise to a **colony** when cultivated under suitable conditions. Therefore, the number of colonies arising from a specific culture volume can be used to calculate the concentration of viable cells within the culture. Use of selective media allows enumeration of specific groups of bacteria. The two most common methods of performing viable counts are the **spread plate** and the **pour plate**.

For pour plates, the culture aliquot is added directly to molten agar (45-50º C), mixed, and poured into an empty petri dish. Colonies arise throughout the medium, so growth of aerobes can be limited. For spread plates, a small volume of cell suspension is dispensed on an agar plate and spread with a sterile glass rod.

It must be recognized that these methods assume that colonies are well-separated and growth conditions are good. Since less than optimal conditions (e.g., cells clumping or inhibition of growth) can skew viable counts, results are expressed as **colony forming units/ml** (CFU/ml), rather than as cells/ml. For statistical accuracy, only plates with between 30 and 300 colonies are counted. An additional limitation of spread plates is that no greater than 0.2 ml is spread on a

plate or the surface will be too wet, causing colonies to run into each other. To meet these requirements, cultures often need to be diluted.

Purpose

To perform viable counts of a yeast culture, using pour and spread plates.

Materials and Methods

Saccharomyces cerevisiae
empty petri plates
sterile 1 ml pipettes
diluent tubes (9 ml 0.7% saline)
tubes of molten PDA agar (9 ml)
PDA plates

1. Dilute yeast cultures 1/10 and 1/100 (9 ml blanks)

2. **Spread Plates**: Dispense 0.1 ml of each dilution to a PDA plate and spread with sterile glass "hockey stick".

3. **Pour Plates**: Add 1 ml of each dilution to 9 ml molten agar, mix, pour into empty petri dishes. Pour and mix in an additional 9 ml molten agar.

4. Incubate plates at 30º C.

Results

Colonies on each plate are to be counted. Spread plates represent 10^{-2} and 10^{-3} dilutions, while pour plates are 10^{-1} and 10^{-2} dilutions.

MPN Enumeration of Mesophiles and Thermophiles in Soil

The enumeration of bacteria is a basic laboratory skill applied to many investigations (*e.g.,* does compound X increase/inhibit growth, which germicide is most effective, how rapidly is the culture growing/dying) in a variety of methods. Cell counts can be direct, where the cells themselves are counted, or indirect, where a property of the cells is measured. Both direct or indirect counts can be made for the total number of cells or for only the viable cells. The most common viable count is a direct count of colonies growing on spread or pour plates inoculated with a diluted sample. Each colony is assumed to arise from a single cell so the number of colonies represent the number of live cells per inoculum volume. This is a very good approach to enumerating viable cells and has the benefit of providing isolated colonies.

Another, somewhat less common, method of making a viable count is the Most Probable Number (MPN) method. Like plate counts, the source of organisms is diluted and the dilutions used to inoculate cultures. Unlike plate counts, broth cultures are inoculated, several replicates at each dilution. The number of tubes showing a positive result (*e.g.,* growth, color change, gas production) are used to determine the most probable number based on a statistical distribution. Since direct counts of cells are not made, this can be considered an indirect, viable count.

The MPN method has several advantages over plate counts. In comparison, it requires less medium and is faster to perform. In addition, inoculation in broth is usually less stressful to the cells than inoculation onto a solid medium or into hot medium, so cell counts may be higher. In some instances, a specialized atmosphere (*e.g.,* high levels of CO_2) may be sealed in tubes and not require additional apparatus. The largest disadvantage in comparison to plate counts is that isolating colonies require an additional step.

Think about the MPN in this manner: if we inoculated a series of tubes in 10-fold dilutions and the highest dilution showing growth was 10^{-4}, we would assume that the cell concentration was at least 1×10^4 but less than 1×10^5. If we wanted more precision, we could inoculate several tubes at each dilution. In practice, MPN methods use 3, 5, or 10 replicates per dilution and the results can be interpreted using published tables.

In this experiment, a three-tube MPN method will be used to enumerate the number of mesophilic and thermophilic microbes in two soil samples. Dilutions of each soil sample from 10^{-2} to 10^{-10} will be used to inoculate two sets of three-tubes, one set to be incubated at 25ºC and the other at 55ºC. The proportions of mesophiles and thermophiles will be compared to determine if the soils are different.

Since the data in this experiment is nominal (mutually exclusive categories of mesophilic or thermophilic), comparisons may be made using the chi-square statistic. Recall from the background on the chi-square statistic in *Looking for Relationships* that the null hypothesis in this case would be that soil source has no effect on the concentrations of mesophiles and thermophiles. An alternative hypothesis is that the concentrations of these organisms will vary with their source. Information on how to calculate the chi-square value and a table of critical values is provided in *Looking for Relationships* but for this experiment a spreadsheet has been constructed to simplify the calculation.

As a result of this experiment, you should gain experience using a three-tube MPN for enumeration and see how the application of the chi-square statistic enables determination of whether variables with nominal data are associated or independent.

PROCEDURE

1. Each group of four students will be given 10^{-1} dilutions of two different soils. Perform serial 10-fold dilutions using the tubes containing 9 ml 0.85% saline from 10^{-2} through 10^{-8} for each soil sample.

2. Inoculate six tubes of tryptic soy broth (TSB) from each of the dilutions you made, labeling each with the soil sample and the dilution. Split the six replicate series for each soil into two sets of three tubes.

3. Incubate one set of the three-tube series for each soil at 25ºC (or room temperature - but record the temperature) and the other set for each soil at 55ºC.

4. After at least 48 h incubation, record the number of tubes showing growth at each dilution for each temperature and sample. Using the table below (Table 2), interpret your data and determine the concentration of bacteria. If possible, the highest dilution to use should show no growth in any of the tubes (dilution to extinction).

5. Using the chi-square spreadsheet, determine if the soils are the same or different in regard to concentrations of mesophilid and thermophilic organisms.

MPN EXAMPLES

Table 1. Number of tubes showing growth from three samples at different dilutions.

Sample	Dilutions					
	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
A	3	3	2	1	0	0
B	3	2	2	0	0	0
C	3	3	3	3	3	1

Using the data in Table 1, we see that for Sample A we should use the results from the 10^{-3} - 10^{-5}
dilutions. If we look up the growth pattern, 2-1-0, in the MPN table (Table 2), we get a value of 0.15 cells/ml in the inoculum for the middle tube (the 10^{-4}). If we multiply 0.15 cells/ml by the dilution factor, 10^4, we determine the concentration in Sample A is 1.5 x 10^3 cells/ml.

In Sample B, we use the results from the 10^{-2} - 10^{-4} dilutions, that is 2-2-0. From Table 2, that is

equivalent to 0.21 cells/ml in the 10^{-3} dilution or a concentration of 2.1×10^{2} cells/ml in Sample B.

In Sample C, we do not have dilution to extinction so we use our three highest dilutions, 10^{-4} - 10^{-6}. From these data (3-3-1) and Table 2, we calculate a concentration of 4.6×10^{5} cells/ml.

VALUES FOR 3-TUBE MPN

Table 2. Values for a Three-Tube MPN assay (Lindquist, J. http://www.jlindquist.net/generalmicro/102dil3.html)

No. of Positive Tubes in			MPN in 2nd Set		No. of Positive Tubes in			MPN in 2nd Set
1st Set	**2nd Set**	**3rd Set**			**1st Set**	**2nd Set**	**3rd Set**	
0	0	0	< 0.03		2	0	0	0.091
0	0	1	0.03		2	0	1	0.14
0	0	2	0.06		2	0	2	0.20
0	0	3	0.09		2	0	3	0.26
0	1	0	0.03		2	1	0	0.15
0	1	1	0.061		2	1	1	0.20
0	1	2	0.092		2	1	2	0.27
0	1	3	0.12		2	1	3	0.34
0	2	0	0.062		2	2	0	0.21
0	2	1	0.093		2	2	1	0.28
0	2	2	0.12		2	2	2	0.35
0	2	3	0.16		2	2	3	0.42
0	3	0	0.094		2	3	0	0.29
0	3	1	0.13		2	3	1	0.36
0	3	2	0.16		2	3	2	0.44
0	3	3	0.19		2	3	3	0.53
1	0	0	0.036		3	0	0	0.23
1	0	1	0.072		3	0	1	0.39
1	0	2	0.11		3	0	2	0.64
1	0	3	0.15		3	0	3	0.95
1	1	0	0.073		3	1	0	0.43

1	1	1	0.11		3	1	1	0.75
1	1	2	0.15		3	1	2	1.2
1	1	3	0.19		3	1	3	1.6
1	2	0	0.11		3	2	0	0.93
1	2	1	0.15		3	2	1	1.5
1	2	2	0.20		3	2	2	2.1
1	2	3	0.24		3	2	3	2.9
1	3	0	0.16		3	3	0	2.4
1	3	1	0.20		3	3	1	4.6
1	3	2	0.24		3	3	2	11
1	3	3	0.29		3	3	3	>24

Use the chi-square spreadsheet (online) to compare your results. Fill out the worksheet on the next page.

SOIL MESOPHILES AND THERMOPHILES WORKSHEET

Soil Sample A

	Number of Tubes Showing Growth									
Temp	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}	10^{-10}

Soil Sample B

	Number of Tubes Showing Growth									
Temp	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}	10^{-10}

MPN Values

Temp	Sample A	Sample B

***P* value =**

In regard to the proportions of mesophiles and thermophiles, are the soils the same or different? Explain.

9 PIPETTING AND SPECTROPHOTOMETRY

Background

The pipet is a fundamental tool in biological and chemical laboratories and its proper use is an important basic skill. Although the principle behind using pipets is simple, it is critical that sufficient practice be performed to develop reproducibility. Serological pipets, the most common pipets we use, are easy to read by holding the pipet vertically and observing the graduation lined up with the bottom of the meniscus. Although that appears simple, there are a number of factors that need to be considered. For example, what is the volume of the pipet? Although 1, 5, and 10 ml pipets are most common, other sizes, larger and smaller, can be purchased. Knowing the volume of the pipet being used is related to observing what volumes are represented by the graduations. Another point to consider is if the pipet is "to deliver" (marked T.D. or has a double ring at the top) or "to contain" (marked T.C. at the top). Pipets that are T.D. deliver the desired volume when fluid is allowed to flow freely from the pipet; T.C. pipets release the chosen volume when the fluid is released to a certain point. (In other words, T.D. pipets include the volume in the tip while T.C. pipets do not.)

Micropipets are small hand pumps that allow the user to set a volume (typically 1-1000 µl) and have disposable tips. Depressing a spring-loaded plunger evacuates the pump, allowing the preset volume to be sucked into the pipet tip. Depressing the plunger once more dispenses the fluid. In order to totally evacuate the pipet tip, the plunger is actually depressed past the point used to fill the pump. Although simple to use, variations between samples can occur due to slight differences in the amount the plunger is depressed or sloppiness in setting the volume. (Dropping the micropipet can also wreak havoc with its calibration.) The more practice one has with a micropipet, the more reproducible the volume transferred.

There are many other pipets used in the lab, for example glass Pasteur pipets and disposable plastic pipets with an integrated bulb (often sterile and packaged individually). These types of pipets are useful for transferring solutions or even inoculating cultures but do not provide the quantitative precision of serological or micropipets.

In this exercise, pipetting precision and reproducibility will be investigated by performing replicate dilutions of methylene blue and measuring their absorbance with a spectrophotometer. Determination of accuracy would require a value from a known. Since we don't have this value, we will use the Central Limit Theory for its determination. Furthermore, dilutions of a yeast suspension will be used to observe the relationship between cell density and absorbance.

STATISTICS BACKGROUND

Statistics is a mathematical field that allows us to summarize and analyze data. The two major subdivisions are **descriptive statistics** and **inferential statistics**. In descriptive statistics, the main goal is data reduction. That is, large amounts of data are summarized, typically with graphs or tables and by calculating measures of central tendency and variability. Although these summaries are convenient for simplifying data and making patterns more obvious, it should be recognized that much of the detail is lost during summarizing. Descriptive statistics are employed when the whole population can

be sampled.

When the whole population cannot be sampled (the typical situation), a sample (or several) is withdrawn from the population and inferences about the population are made from the characteristics of the sample. Thus, it is assumed that a sample is a good representative of the population. However, this is not always the case and statistics also allows us to decrease our uncertainty about drawing conclusions from incomplete data.

Measures of Central Tendency

A **distribution** is created when data variables are arranged in order from highest to lowest, typically represented graphically as a **frequency distribution**, plotting the value of the variable versus the frequency of that value's occurrence. It was shown quite a long time ago that if the number of data points is very large, the distribution becomes a normal distribution. Researchers can learn a lot from these plots, including the shape of the distribution, the range of values, and the most common value. Measures of **central tendency**, that is the values that the distribution seems to cluster around, are some of the most commonly calculated statistics.

The **mean** is one of the most used statistics in all manners of research. It is the arithmetic average of values, calculated as the sum of all values divided by the number of values. Although the mean provides a simple summary of a distribution, it doesn't indicate anything about the range of values. The distribution may have a very tight distribution around the mean or may be so spread out that a peak is hard to identify. Another shortcoming of the mean is that it is sensitive to extreme values. A single outlying data point can skew the distribution so that the mean no longer represents the peak of the curve. Because of this, other descriptors of the central tendency may actually be more useful.

The **mode** is the most frequently occurring value. It's where the distribution displays a peak (or peaks, in the case of a multi-modal distribution). It is the only descriptor of central tendency possible with nominal data. (Nominal data classifies items into mutually exclusive groups and can only be classified as equal or not equal, for example, "male" or "female.")

The **median** is the midpoint of the distribution, with half of the values on either side of it. Another way to say this is that the median represents the 50th percentile. The median can tell us about the shape of the distribution. In a normal distribution, the mean and median (and mode) are the same. If the distribution is skewed, the mean is closer to the mode than is the median.

The Central Limit Theorem

The **Central Limit Theorem** states that the means of independently drawn random samples will be approximately normally distributed, if the sample size is large enough. In other words, even if a distribution is not normal, the distribution of the means will be. Means calculated from observation sets containing thirty sample values is considered large enough to produce a normal distribution of means from non-normally distributed data, but 10 sample values per observation set may be enough in many cases. The mean of a sampling distribution of the means is considered the same as the

population mean and is often called the **expected value**. When we calculate a mean from random samples drawn from a population, we expect that mean to be the same as the population mean. The dispersion (standard deviation) of the sampling distribution of the means is called the **standard error** because it is drawn from a sampling distribution rather than a data distribution. It represents how confident we should be that a sample mean represents the population mean. The standard error is equivalent to the population standard deviation divided by the square root of the number of samples in each observation set. The standard error is smaller than the population standard deviation so distribution of the means has less dispersion than the distribution of the population. Thus, a sampling distribution of the means can provide more accurate inferences about a population than a distribution from a large pool of data.

In practice, sampling distributions of the mean are usually not constructed but knowledge of the Central Limit Theorem increases our confidence when making an inference based on a single sampling. We know that if we take several sets of samples from a population, they would have a normal distribution. So, a mean from a single set of samples taken from a population represents one mean in a normal distribution. Based on the Central Limit Theorem it is possible to calculate the probability that a sample or predicted outcome is significantly different from the population mean, but that's a discussion for another time.

Hypothesis Testing

We know that a hypothesis is a testable prediction that is consistent with specific observations. With testing, it is possible to show that a hypothesis is false but not true. At best, we refuse to reject a hypothesis based on current data but recognize that as yet unexamined conditions may show the hypothesis to be false sometime in the future. In statistical hypothesis testing, it is common to reduce the question at hand to two outcomes (*e.g.,* A and B are the same, A and B are not the same). We can state these outcomes as the **null hypothesis** (**H_o**), or the hypothesis of no difference, and the **alternative hypothesis** (**H_a**).

Two of the most common statistical tests used for determining if a significant difference exists between two or more samples is the *t*-test and ANOVA (analysis of variance). The *t*-test is a powerful and robust test for determining if two populations have different means when samples are independent and random and the measured variable is continuous and normally distributed. ANOVA is more appropriate (and less error prone) than running multiple *t*-tests when there are more than two samples to compare. In ANOVA there are two types of variance to consider: error variance (within-groups variance) and treatment variance (between-groups variance). When samples are independent and random, the within-groups variance should be the same. Therefore, any difference between the variances of the measurements would be due to the treatment. ANOVA can indicate a difference between the treatments but additional tests (*post hoc* texts) are necessary to find which treatments are significantly different. A level of significance must be selected for statistical hypothesis testing, designated as an alpha level. An alpha value of 0.05 is the level usually selected as it offers a good compromise to avoid Type I (rejecting a null hypothesis that is true) or Type II (not rejecting a null

hypothesis that is false) errors.

Materials

serological pipets (1 and 10 ml)
13 x 100 mm test tubes
spectrophotometers
microscopes
hemocytometers
methylene blue solution (1 mg/ml), 25 ml per group of 4 students
dense yeast suspension in 5% NaCl, 100 ml per group of 4 students

Procedure

Determination of Variability in Pipetting

1. Dilute the methylene blue solution 1/400 using 1 ml serological pipets with solutions containing dye and 10 ml serological pipets for the diluent (use water). **Each person should do this at least five (5) times.** You may use any dilution scheme desired but each person in the group should use the same scheme. For example, I would do three dilutions: 1/4 → 1/10 → 1/10.
2. Make five more 1/400 dilutions of the methylene blue dye **using only the 10 ml serological pipets**. Each person should **make at least five dilutions** this way.
3. Read the absorbance of each 1/400 dilution at 600 nm and record the results.

Determining the Relationship Between Cell Density and Absorbance

1. Each group needs to dilute the yeast suspension in water to 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90% of the stock.
2. Measure the absorbance of each dilution at 660 nm. **Use the same tube for each reading unless we have brand new tubes.** What fluid will you use in your blank?
3. Measure the absorbance of each dilution at 440 nm.
4. Plot the data and **if a linear portion is not apparent, perform additional dilutions and measure their absorbance.**
5. Use a hemocytometer to make a total cell count of the stock yeast suspension. (See Chapter 7.)

Results

1. Record your data.
2. Analyze the dilution data to see if there are any differences between the absorbencies resulting from using different pipets. Perform a *t*-test on your data (two sets of five measurements) to determine if there is a significant difference due to the pipet used. Perform ANOVA tests between the measurements taken by all four members of your group. Make two comparisons, one between the data with the 1 ml pipets and

another between the data from the 10 ml pipets.

3. Record the means and standard deviation of your data. (*t*-test table)
4. Plot the data (A_{660} vs. Relative Density) from the yeast dilutions and determine the range of cell density where absorbance is linear. Do the same with A_{440} data.
5. Once you receive the data from the whole class, plot a frequency distribution of all the data and a frequency distribution of the means from each individual's data set. Calculate the mean and standard deviation for the complete class data and the mean and standard error for the class means.

Conclusions

1. Are dilutions made using different pipets the same? Consider your measurements and the measurements of taken by your group. Explain how you made your decision.
2. Based on the Central Limit Theory, what should the absorbance value of the dilutions be?
3. If you answer to question 1 was "no," which dilution method appears most accurate and precise?
4. At what absorbance did cell density become nonlinear? Was there a difference due to the wavelength used?
5. Using the total cell count, what is the relationship between cell concentration and absorbance?

Pipetting and Spectrophotometry Lab Worksheet

Names:

Results and Discussion

Part 1: Diluting methylene blue

Student:								
	1 ml	**10 ml**	**1 ml**	**10 ml**	**1 ml**	**10 ml**	**1 ml**	**10 ml**
Mean								
Std. Dev.								
***t*-test *P* value**								
ANOVA *P*-value, 1 ml								
ANOVA *P*-value, 10 ml								

1. Which pipet gives more precise results? Explain.

2. Which pipet gives more accurate results? Explain.

3. Is there a difference between using the 1 ml or 10 ml pipets? Explain.

4. Is there a difference between who is using the pipet? Explain.

Part 2: Measuring absorbance of a yeast suspension

1. Attach plots of relative density of the yeast suspension versus absorbance at 440 and 660 nm.

2. At what absorbance did cell density become nonlinear? Was there a difference due to the wavelength used?

10 STATISTICS WITH A HEMACYTOMETER

Measures of Central Tendency

Overview

This exercise incorporates several different statistical analyses. Data gathered from cell counts with a hemacytometer is used to explore frequency distributions of data, measurements of central tendencies, the Central Limit Theorem, and hypothesis testing with ANOVA.

Objectives

- To learn to calculate specific statistics (*mean, mode, median*) for summarizing the center of a data distribution.
- To plot data distributions.
- To investigate the Central Limit Theorem by plotting the frequency distribution of the means, calculating the mean and standard error, and comparing them with similar treatments of the pool of sample data from the entire class.
- To test the null hypothesis that there is no significant difference between cell counts made from one, five, or twenty-five hemacytometer grids.

Procedure

1. Using the yeast cell suspension provided, make 10 counts with a hemacytometer. Record your data as shown in Figure 1; record the count in the center grid, the four corners, and the remaining twenty grids for each of the 10 samples. Transfer your data to a spreadsheet. Name the worksheet "Personal Data" by right clicking on the name of the worksheet, selecting "Rename" from the pop-up menu, typing "Personal Data" and hitting the "Enter" key.

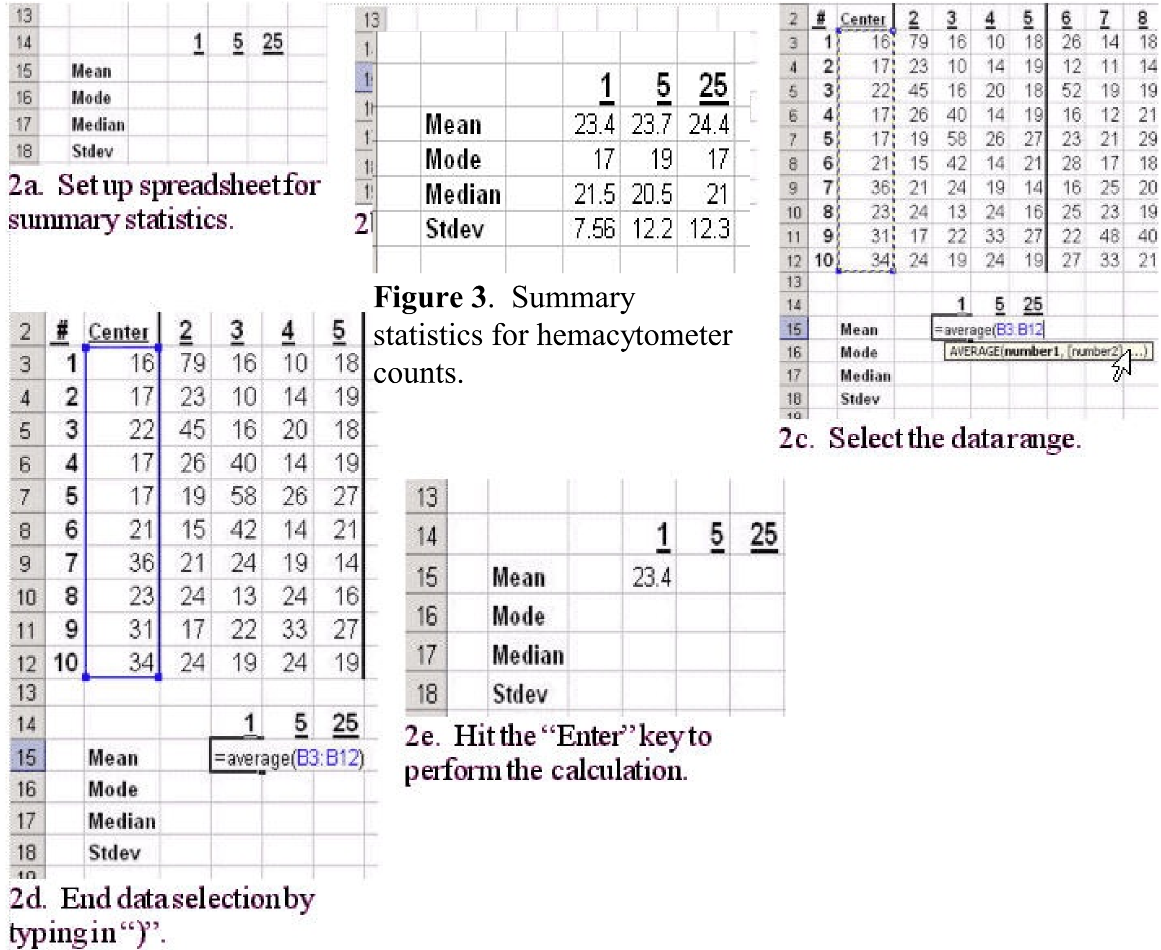
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1			Corners																							
2	**#**	**Center**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**
3	**1**	16	79	16	10	18	26	14	18	48	36	19	8	16	39	34	19	13	34	22	21	17	19	23	23	7
4	**2**	17	23	10	14	19	12	11	14	16	14	13	16	14	17	18	17	25	22	17	17	11	19	22	22	14
5	**3**	22	45	16	20	18	52	19	19	22	15	18	31	20	19	20	27	16	38	15	14	16	23	31	27	26
6	**4**	17	26	40	14	19	16	12	21	22	21	21	14	16	23	21	25	29	17	24	37	17	17	32	35	23
7	**5**	17	19	58	26	27	23	21	29	41	32	20	24	44	26	13	27	16	39	26	21	16	13	19	22	32
8	**6**	21	15	42	14	21	28	17	18	22	18	17	46	15	20	17	21	12	12	25	20	14	16	16	22	54
9	**7**	36	21	24	19	14	16	25	20	20	15	21	20	17	21	41	25	18	25	14	33	29	26	27	17	24
10	**8**	23	24	13	24	16	25	23	19	28	18	27	24	12	14	19	22	39	28	18	18	35	22	36	13	42
11	**9**	31	17	22	33	27	22	48	40	25	121	28	20	74	23	40	49	18	30	7	58	45	29	35	40	32
12	**10**	34	24	19	24	19	27	33	21	21	34	43	22	13	24	25	60	33	29	17	25	32	16	29	26	28

Figure 1. Spreadsheet format for the hemacytometer counts data.

2. Calculate the mean, mode, median, and standard deviation. Most spreadsheets have built in statistical functions. In Microsoft® Excel, functions are selected by typing in an equal sign, "=", followed by the function name. The function name for mean, mode, median, and standard deviation are "average", "mode", "median", and "stdev", respectively. For example, to calculate the mean of cell counts from the

center cell, you would type in “=average(” (without the quotation marks), select the range of cells by dragging the mouse over them, and hitting the close parentheses, “)” to select those values. Hitting “Enter” calculates the value. (Figure 2)

3. Repeat this operation for all of the functions with the data for the center grid, the



	1	5	25
Mean	23.4	23.7	24.4
Mode	17	19	17
Median	21.5	20.5	21
Stdev	7.56	12.2	12.3

Figure 3. Summary statistics for hemacytometer counts.

#	Center	2	3	4	5	6	7	8
1	16	79	16	10	18	26	14	18
2	17	23	10	14	19	12	11	14
3	22	45	16	20	18	52	19	19
4	17	26	40	14	19	16	12	21
5	17	19	58	26	27	23	21	29
6	21	15	42	14	21	28	17	18
7	36	21	24	19	14	16	25	20
8	23	24	13	24	16	25	23	19
9	31	17	22	33	27	22	48	40
10	34	24	19	24	19	27	33	21

Figure 2. Steps taken to calculate values using built-in functions.

center grid plus the corner grids, and all twenty five grids (Figure 3).

4. **Calculate a frequency table of your data.**
- First, name a worksheet “Frequency.”
- To get a frequency distribution, there needs to be a data set and bins, the divisions used to group the data. For example, test scores for the class could be grouped into the bins 59, 69, 79, 89. This would give the number of scores that were 59 and below, 60-69, 70-79, 80-89, and above 89 (even though another bin wasn’t listed). Unfortunately, this is not a straightforward, intuitive operation.
- Go to the “Frequency” worksheet and make a list of bins. Make sure that the values exceed the highest value from the count data. Type a value of 0 in cell A1. Select the whole column by clicking on the “A” at the top of the column. Click on“Edit” in the toolbar at the top of the page and select “Fill.” Select “Series” from the pop-

up menu. The default values in the “Series” pop-up menu should be appropriate (Figure 4); enter “100” for the “Stop value.” Click on “OK” and column A should be filled with values from 0-100 with an increment of 1. This is the bin.

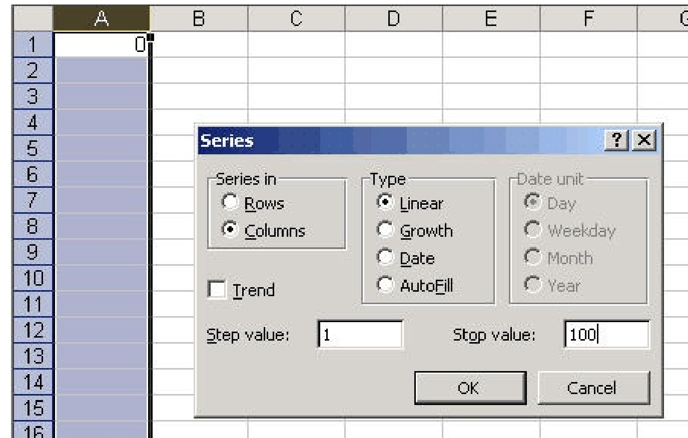
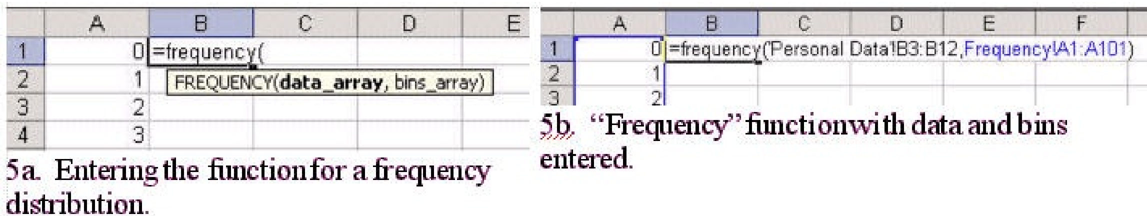


Figure 4. Entering values in a series.

- In cell B1, type “=frequency(” (Figure 5a). Select the data array by returning to the “Personal Data” worksheet (click on the tab), left click on the first value (center grid data, cell B3), and drag the mouse over the remaining values in that column. Type a comma and return to the “Frequency” worksheet. Click on cell A1 and drag the mouse to include all the bin data. Type the close parentheses to end the function (Figure 5b) and hit the “Enter” key. The value in cell B1 should be 0. Select cell B1 and drag the mouse down column B to one cell past the end of the bin list. Hit the “F2” key followed by “Ctrl + Shift + Enter.” This will result in the frequency of occurrence of each value in the bin (column A) appearing next to it in column B. These frequency values represent the data from counting the center grid.



5a. Entering the function for a frequency distribution.

5b. “Frequency” function with data and bins entered.

Figure 5. Using the “Frequency” function.

- Generate frequency data for counting five grids and twenty-five grids by following the same procedure used for the center grid counts. This should result in four columns on the “Frequency” worksheet where column A is the bins, column B is the frequency of counts from the center grid, column C is frequency of counts from the center grid and four corners, and column D is frequency of counts from all twenty-five grids. (You can pretty this up by inserting labels above the columns. Left click in a cell in row 1. Right click and select “Insert . . . ” from the pop-up menu. Choose “Entire row” and click on “OK.” Type the appropriate labels above each column (Figure 6).)

	A	B	C	D
1	Bins	Center	5 Grids	25 Grids
2	0	0	0	0
3	1	0	0	0
4	2	0	0	0
5	3	0	0	0

Figure 11. Labeled columns for frequency data.

5. **Plot a frequency distribution of your data.**

- Use the data on the "Frequency" worksheet to create plots showing the frequency distributions of the data. The "Bins" will be the x-axis and the frequency data will be the data series used.
- Select the columns with the frequency data by left clicking on the column B heading and dragging the pointer to highlight columns C and D, as well.
- Select the "Chart Wizard" from the toolbar (Figure 7).

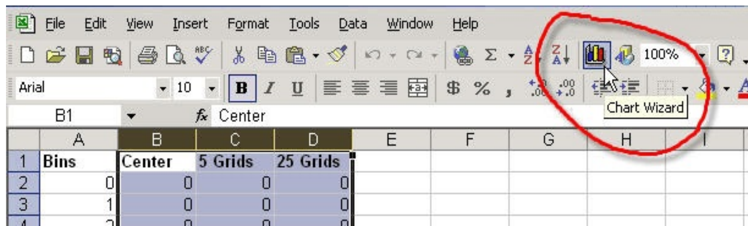


Figure 12. Selecting data to plot and the "Chart Wizard."

- On the first "Chart Wizard" menu ("Chart Type"), select "Area" for the chart type and "Area" for the sub-type (Figure 8). Click on the "Next" button. Click on the "Next" button on the following page ("Step 2 of 4 - Chart Source Data").
- On the next menu page ("Step 3 of 4 - Chart Options"), fill in the information to

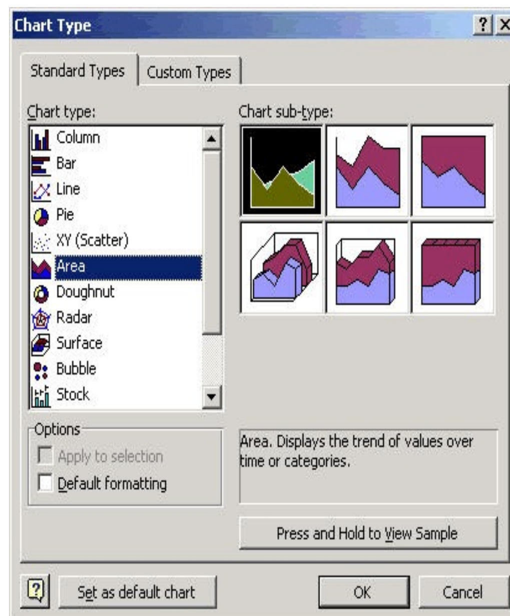


Figure 13. Selecting the chart type.

label the chart and the axes (Figure 9). The labels should appear on the image of the chart. Click on the "Next" button.

- On the final menu page, "Chart Location," make sure that "As object in:" is selected and that the "Frequency" worksheet is listed. Click on the "Finish" button.
- The plot should appear in the "Frequency" worksheet. Feel free to move or enlarge it.
- One of the problems you may encounter is that the third series you selected (all 25 grids) may produce a plot that blocks out the other plots. In order to move the plot to the back, allowing the others to be seen, move the mouse arrow over the plot and click the right hand button. Select "Format Data Series ..." from the pop-up menu. On the "Format Data Series ..." menu (Figure 10), select the "Series Order" menu. Plots on the graph are in the order listed under "Series order:". In this example (Figure 10), the plot made from the center grid counts is in t he back and the plot from counts of all 25 grids is in the front. Move the plot from the 25 grids to the back of the graph by selecting it from the list and using the "Move Up" button to bring it to the back. Move the plot from the 5 grids in the same manner so that all three plots are visible. Click on the "OK" button when you're finished.

6. **Use the class data in the same way that you used your personal data**. That is:

- **Calculate mean, mode, median, and standard deviation** for the entire collection of data.
- **Plot a frequency distribution**. If you can, put it on the same plot as your personal data.
- **Calculate the mean, mode, median, and standard deviation (standard error) from the class means** (for all three data sets: 1 grid, 5 grids, and 25 grids). Remember, in this case the standard deviation is referred to as the standard error because we're the distribution is made up of means and not samples. The standard error should be about the same as the population standard deviation divided by the square root of n (the number of samples for each mean). In this case, n=10 for counts with 1 grid, n=50 for counts with 5 grids, and n=250 for counts with 25 grids.
- **Plot the frequency distribution of the mean for the means from each of the three data sets.**

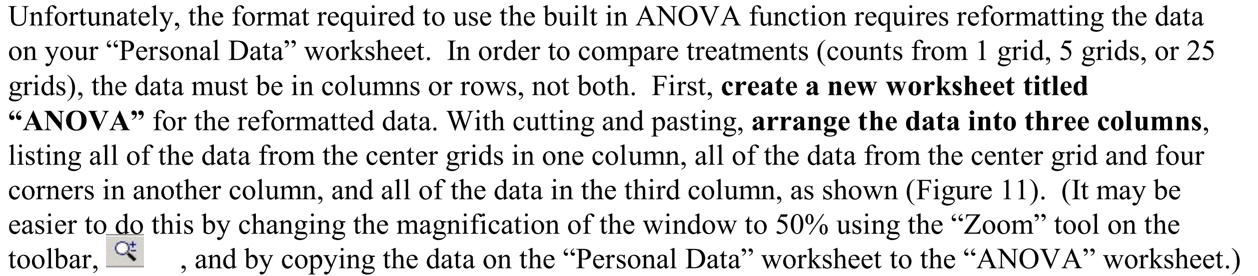
Hypothesis Testing with Data from Hemacytometer Counts

Objective

To determine if there is a significant difference between cell counts made from one, five, or twenty-five grids in a hemacytometer. Since this requires comparing three treatments with data that are random, independent, and where the variance can be assumed to be the same, ANOVA is the appropriate statistic.

Entering and Formatting Data

Microsoft® Excel contains built in functions that simplify calculating statistics.

Unfortunately, the format required to use the built in ANOVA function requires reformatting the data on your "Personal Data" worksheet. In order to compare treatments (counts from 1 grid, 5 grids, or 25 grids), the data must be in columns or rows, not both. First, **create a new worksheet titled "ANOVA"** for the reformatted data. With cutting and pasting, **arrange the data into three columns**, listing all of the data from the center grids in one column, all of the data from the center grid and four corners in another column, and all of the data in the third column, as shown (Figure 11). (It may be easier to do this by changing the magnification of the window to 50% using the "Zoom" tool on the toolbar,  , and by copying the data on the "Personal Data" worksheet to the "ANOVA" worksheet.)

The statistical functions can be found by **clicking on "Tools"** on the toolbar and **selecting "Data Analysis"** (Fig. 12).

If the "Data Analysis" selection is not listed it means that the functions haven't been installed. To install them, you must have access to the installation program, either on CD or through

	A	B	C
1	**1**	**5**	**25**
2	16	16	16
3	17	17	17
4	22	22	22
5	17	17	17
6	17	17	17
7	21	21	21
8	36	36	36
9	23	23	23
10	31	31	31
11	34	34	34
12		79	79
13		23	23
14		45	45
15		26	26
16		19	19
17		15	15
18		21	21
19		24	24
20		17	17
21		24	24
22		16	16
23		10	10
24		16	16
25		40	40
26		58	58
27		42	42
28		24	24
29		13	13
30		22	22
31		19	19
32		10	10
33		14	14
34		20	20
35		14	14
36		26	26
37		14	14
38		19	19
39		24	24
40		33	33
41		24	24
42		18	18
43		19	19
44		18	18
45		19	19
46		27	27
47		21	21
48		14	14
49		16	16
50		27	27
51		19	19
52			26
53			12
54			52
55			16
56			23
57			28

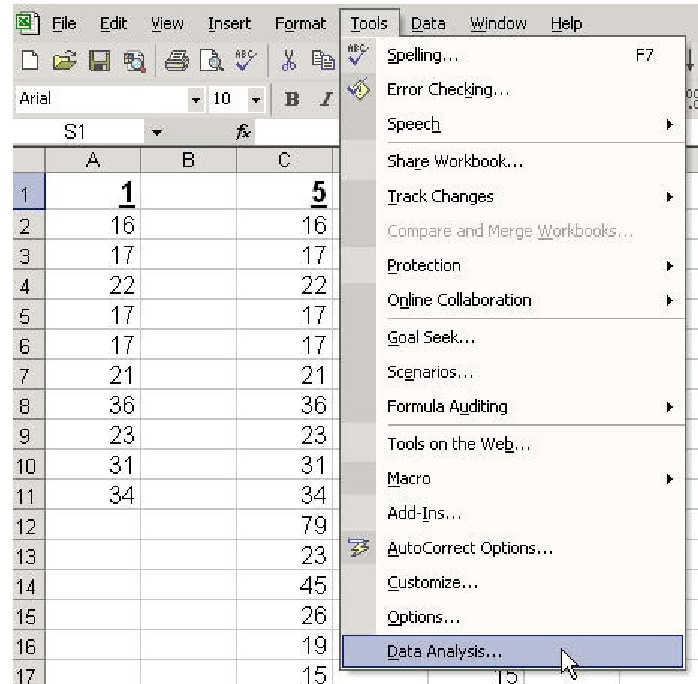


Figure 2. Selecting the "Data Analysis" menu.

a network.

- Select "**Add-Ins**" under the "**Tools**" menu (Fig. 2).
- When the "Add-Ins" menu comes up, choose "**Analysis ToolPak**" (Fig. 3) and click on "OK."

When you select "Data Analysis" on the "Tools" menu (Fig. 2), the "Data Analysis" menu pops up. Choose "**ANOVA: Single Factor**" (Fig.4).

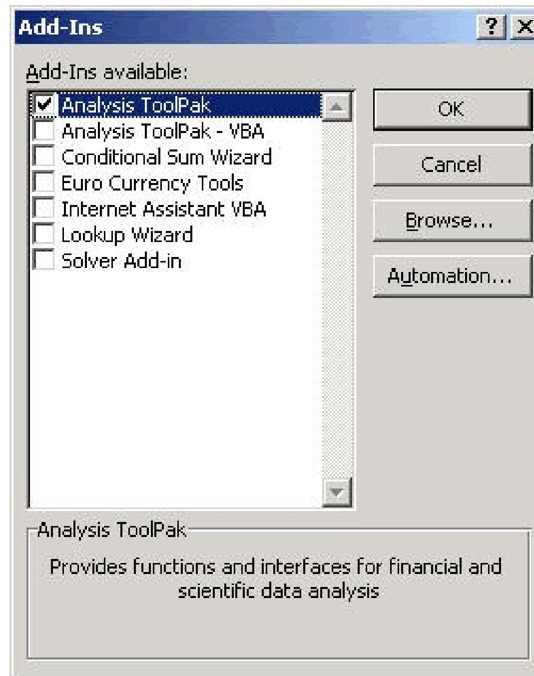


Figure 3. Adding analysis tools, including statistical functions.

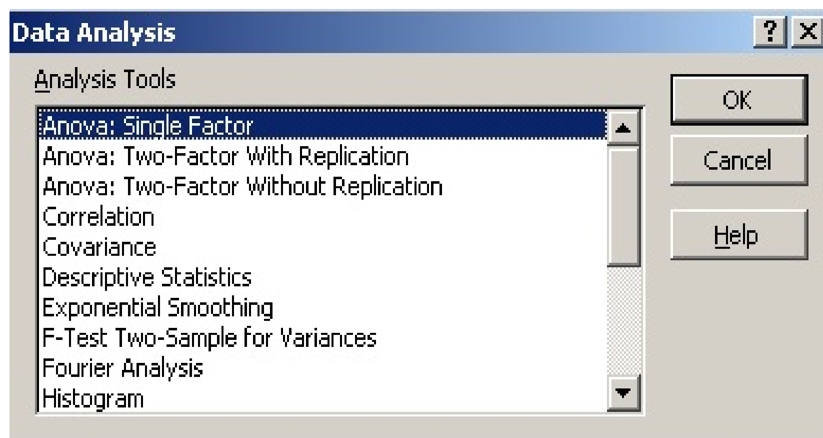


Figure 4. Select "Anova: Single Factor" from the "Data Analysis" menu.

Complete the "Anova: Single Factor" pop-up menu as follows:

- For "**Input Range**:" select the **data in all three columns**, including the headings in Row 1.
- Make sure that "**Columns**" is selected in "**Grouped By:**."
- Select "**Labels in First Row**."

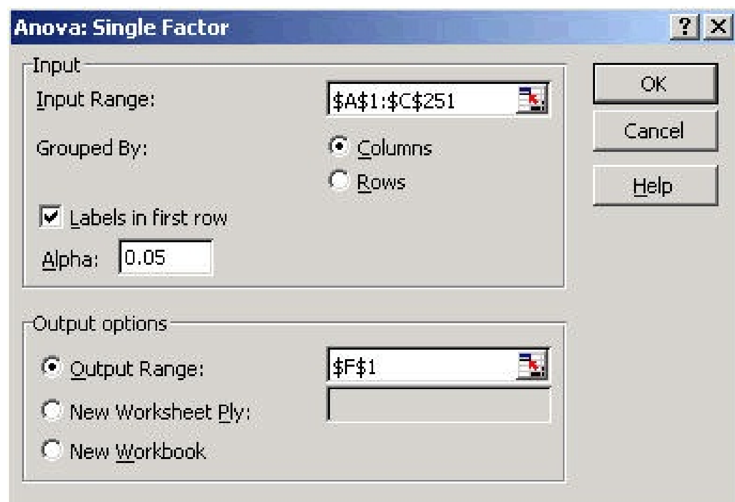


Figure 5. Data entered for ANOVA analysis.

F	G	H	I	J	K	L
Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
1	10	234	23.4	57.15556		
5	50	1183	23.66	149.331		
25	250	6093	24.372	152.0257		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	28.16955	2	14.08477	0.094647	0.909721	3.02515
Within Groups	45686.02	307	148.8144			
Total	45714.19	309				

Figure 6. ANOVA results from personal data.

- "**Alpha:**" should be 0.05.
- For "**Output Range**" select a cell near the data. (In this example, F1.)
- When complete, the "ANOVA: Single Factor" menu should look like Figure 5. Click "OK" and a table will be generated, starting with cell F1 (Fig. 6).

Compare the calculated *F* value (0.0.094647, in this example) with the **critical *F* value** (*F crit*)(3.02515). Since the *F* value is less than the critical *F* value, we cannot reject the null hypothesis that there is no difference between cell counts made by counting one grid, five grids, or twenty-five grids of the hemacytometer. (Note how much larger the *P*-value is than the alpha of 0.05.)

Repeat this using the class data.

APPENDIX

1. **Colony Morphology**

2. **Using Microsoft® Excel to Plot and Analyze Kinetic Data**

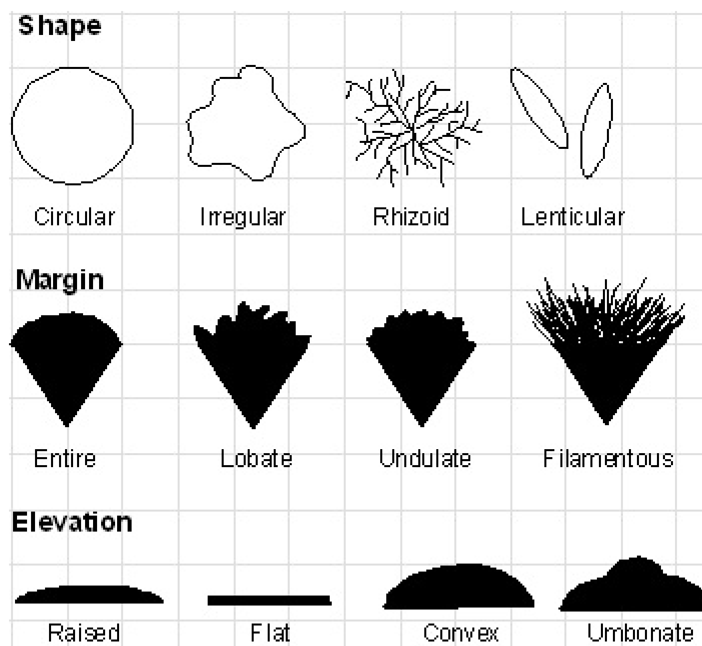
3. **Using Microsoft® Excel to Analyze Data from the Disk Diffusion Assay**

4. **Looking for Relationships**

Appendix 1. Colony Morphology

Isolated colonies should be described using common terms relating to six different features. These features are:

- **Size**
  - diameter (mm) after specific incubation time
  - pinpoint, small, medium, or large (not as useful)
- **Colony surface**
  - smooth or rough
  - shiny or dull
  - firm, viscous, or dry
- **Pigmentation**
  - any color, soluble or insoluble
  - opaque, translucent, iridescent, pr opalescent
- **Shape**
  - circular, irregular, rhizoid, or lenticular (lenticular usually applies only to embedded colonies)
- **Edge or Margin**
  - entire, lobate, undulate, or filamentous
- **Elevation**
  - raised, flat, convex, or umbonate



Appendix 2: Analyzing Pipetting Data

Entering and Formatting Data

Open Excel. Set up the spreadsheet page (Sheet 1) so that anyone who reads it will understand the page (Figure 1).

- Type a **title** in the cell in the upper lefthand corner, cell **A1**
- **Label column A, B, C, and D in Row 3** with the members of your group.
- **Label each column beneath a student's name with "1ml"** (Row 4)
- Move the cursor down to **Row 11** and **label each column beneath a student's name "10ml."**
- Enter the data (spectrophotometry readings) for each student from each pipette.

	A	B	C	D
1	Pipetting			
2				
3	Student A	Student B	Student C	Student D
4	1ml	1ml	1ml	1ml
5	0.202	0.267	0.316	0.19
6	0.232	0.302	0.321	0.202
7	0.296	0.27	0.327	0.205
8	0.204	0.313	0.308	0.192
9	0.27	0.334	0.316	0.223
10				
11	10 ml	10ml	10ml	10ml
12	0.312	0.28	0.327	0.236
13	0.298	0.337	0.32	0.181
14	0.37	0.279	0.318	0.232
15	0.331	0.296	0.366	0.175
16	0.327	0.294	0.328	0.173

Figure 1. Initial setup

The t-test with Microsoft® Excel

The easiest way to perform **statistical analyses** with a spreadsheet is to **use the built-in functions**. With Excel, they can be found by **clicking on "Tools"** on the toolbar and **selecting "Data Analysis"** (Fig. 2). Scroll down to "**t-Test Two-Sample Assuming Equal Variance**."

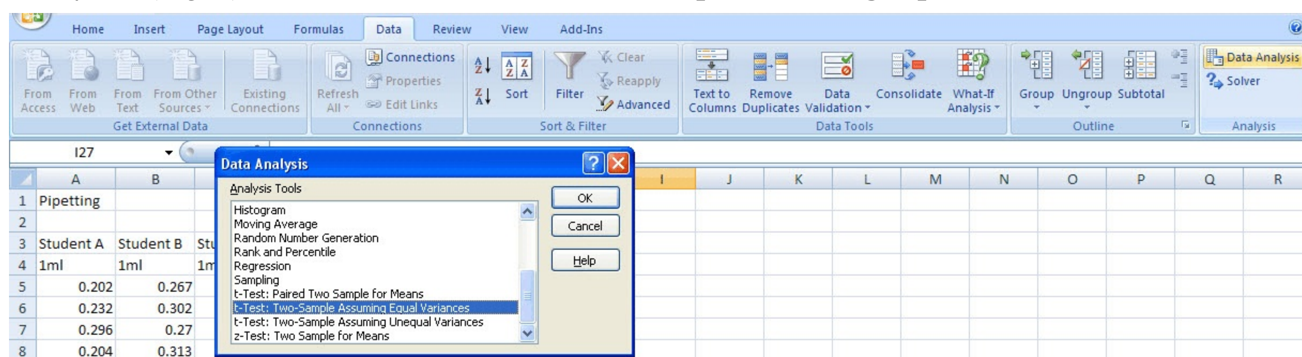


Figure 2. Selecting the *t*-test from the **Data Analysis** menu.

If the "**Data Analysis**" selection is not listed it means that the functions haven't been installed. To install them, you must have access to the **installation program**, either on CD or

through a network.

- Select the "**Office Button**" (the round button with the Microsoft emblem, in the upper left corner)
- Click on "**Excel Options**" on the bottom of the drop-down menu.
- On the "**Excel Options**" menu, select "**Add-Ins**" followed by selection of "**Analysis ToolPak**."
- Click "OK" and the Data Analysis Tools should be installed on the right hand side of the "**Data**" menu tab.

Once you've selected "**t-Test Two-Sample Assuming Equal Variance**" a menu box pops up and you need to select your data (Fig. 3). **Each person chooses their own data for this analysis.** Include the labels "1ml" and "10ml" in your selection and make sure to check the "**Labels**" box on the menu. Once all the data is entered, click "**OK**" and a table of values will appear at the output location you selected (Fig. 4).

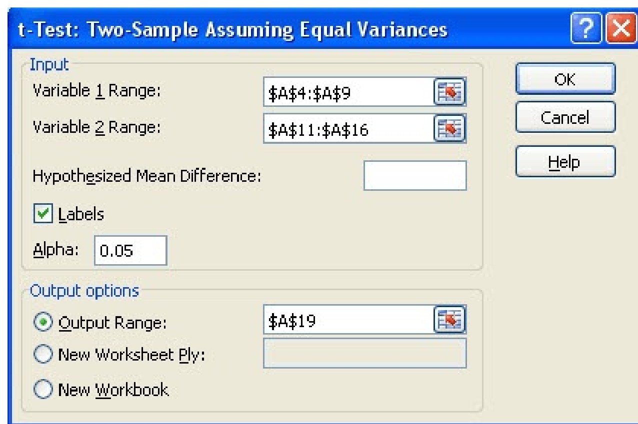


Figure 3. Pop up menu for the *t*-test.

19	t-Test: Two-Sample Assuming Equal Variances		
20			
21		*1ml*	*10 ml*
22	Mean	0.2408	0.3276
23	Variance	0.0017092	0.0007323
24	Observatio	5	5
25	Pooled Vari	0.0012208	
26	Hypothesiz	0	
27	df	8	
28	t Stat	-3.9280442	
29	P(T<=t) one	0.0021843	
30	t Critical on	1.859548	
31	P(T<=t) two	0.0043687	
32	t Critical tw	2.3060041	

Figure 4. Results of the *t*-test.

The *t*-test table provides the mean for each set of samples, as well as the variance. Further down the table are the results of the *t*-test for a one-tailed and a two-tailed test. To test our null hypothesis of no difference between dilutions made with either pipet, we select the two-tailed test. The *P* value from the two-tailed test, 0.0043687, is less than the α value of 0.05 (Fig. 3) so the results are **significant.** In statistics, saying that something is significant means that it is different. So, in this example, we reject the null hypothesis and say that the pipet used does matter.

ANOVA with Microsoft® Excel

When comparing more than two treatments, analysis of variance (ANOVA) is used instead of the *t*-test. ANOVA will be applied for comparisons of the absorbance values from dilutions by the different members of each group, first using the 1ml pipettes and then using the 10ml pipettes. The ANOVA function requires that the data to be analyzed are in contiguous columns or rows.

On the "**DATA**" menu, select "**Data Analysis**." From the drop-down menu, select "**Anova: Single Factor**." For the "**Input Range:**" use the mouse to highlight all of the readings from the dilutions with the 1ml pipette for all four members of the group (Fig. 5). Hit "**Enter**" on the keyboard. Select "**Output Range:**" in a similar manner. Two tables of data is presented (Fig. 6).

	A	B	C	D	E	F	G	H
3	Student A	Student B	Student C	Student D				
4	1ml	1ml	1ml	1ml				
5	0.202	0.267	0.316	0.19				
6	0.232	0.302	0.321	0.202				
7	0.296	0.27	0.327	0.205				
8	0.204	0.313	0.308	0.192				
9	0.27	0.334	0.316	0.223				

Figure 5. Selecting data for ANOVA

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Column 1	5	1.204	0.2408	0.001709
Column 2	5	1.486	0.2972	0.00082
Column 3	5	1.588	0.3176	4.93E-05
Column 4	5	1.012	0.2024	0.000173

ANOVA

Source of Varia	SS	df	MS	F	P-value	F crit
Between (	0.041535	3	0.013845	20.1272	1.12E-05	3.238872
Within Gr	0.011006	16	0.000688			
Total	0.052541	19				

Figure 6. ANOVA output (1ml pipettes)

The ANOVA table contains the information related to the null hypothesis (it doesn't matter who is making the dilution). Since the *P*-value is so small, it suggests that there is a significant difference between the people making the dilution. (Although it does not indicated what those differences are.) Repeat the ANOVA test with the data from using the 10ml pipette.

Determining Accuracy

Accuracy, in this case, is defined as which pipette gave dilutions that were closest to the actual absorbance value for this dilution of methylene blue. We could calculate a value using an appropriate wavelength of light, diameter of the cuvette, concentration of the dye, and the dye's corresponding extinction coefficient. Or, we can rely on the Central Limit Theorem to determine what the value should be. Recall that the Central Limit Theorem says that the mean of a distribution of means drawn from the same population will be the same as the population mean. Additionally, the standard error of the distribution of means will be less than the standard deviation of the population. Although this works best with large sample sizes (> 30), our sample size of 5 should be adequate since the data should be normally distributed. What this means, is that even with a small sample size, the distribution of means should approximate a normal distribution. Thus, the mean of all the means from the class will be the absorbance value we want for our dilutions. Once I have everyone's data, I will post it on my website so that each person can calculate the mean of means to compare with their individual data and determine if dilutions with a 1ml pipette or 10ml pipette were more accurate.

Appendix 3. Using Microsoft® Excel to Plot and Analyze Kinetic Data

Entering and Formatting Data

Open Excel. Set up the spreadsheet page (Sheet 1) so that anyone who reads it will understand the page (Figure 1).

- Type a **title** in the cell in the upper lefthand corner, cell **A1**
- **Label column A** as the substrate concentration in cell **A3**
- **Label column B** as the reaction rate for 30s in cell **B3**
- **Label column C** as the reaction rate for 1min in cell **C3**
- Adjust column widths to fit the labels by clicking on the column heading and dragging the border to the appropriate width

	A	B	C	D
1	Mock Enzyme Kinetics			
2				
3	S (pennies/m^2)	v (pennies/30 s)	v (pennies/min)	
4				

Figure 1

Enter your data pairs in the appropriate columns. (Don't forget to **enter 0,0** for one of your data pairs.) If your data was not collected in order of increasing substrate concentration, enter the data pairs in the order collected and **sort** them in ascending order (Fig. 2).

- **Click and drag** over the cells that contain the data pairs
- Choose **Data > Sort**

	A	B
1	Mock Enzyme Kinetics	
2		
3	S (pennies/m^2)	v (pennies/30 s)
4		
5	75	31
6	15	10
7	10	6
8	60	28
9	30	16
10	10	6
11	13	8
12	130	39
13	50	23
14	10	7
15	25	13
16	30	15
17	18	10
18	13	9
19	15	9
20	100	34
21	40	20
22	18	11
23	22	12
24	0	0

Figure 2

- On the **Sort** menu, select “Column A” in the drop down menu for **Sort by**
- Click “OK” (Fig. 2)

Once the data is sorted in ascending order, the reaction rate for 1min can be calculated in column C by entering the formula **=(B4*2)** in **cell C4**. You can **copy** and **paste** the formula into the other cells in column C by clicking the right-hand button on the mouse and making the appropriate selection (Fig. 3) or by double-clicking and holding on the box in the lower right hand corner and dragging the mouse down the column for the length of the data.

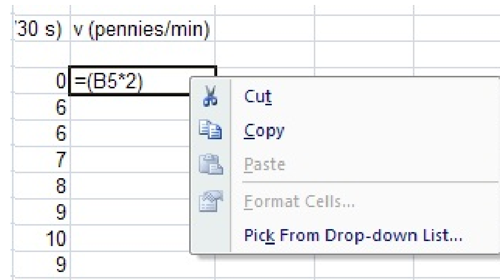


Figure 3

For now, skip column D and **label row 3 in columns E and F “1/S” and “1/v,”** respectively. Calculate the values for these columns by taking the inverse of the values in column A and column C (*e.g.*, skip formulas for values of “0,” so in cell E6 type **=(1/A6)** and in cell F6 type **=(1/C6)**). Copy and paste the formulas into the other cells (Fig. 4).

	A	B	C	D	E	F
1	Mock Enzyme Kinetics					
2						
3	S (pennies/m^2)	v (pennies/30 s)	v (pennies/min)		1/S	1/v
4						
5	0	0	0			
6	10	6	12		0.1	0.083333
7	10	6	12		0.1	0.083333
8	10	7	14		0.1	0.071429
9	13	8	16		0.076923	0.0625
10	13	9	18		0.076923	0.055556
11	15	10	20		0.066667	0.05
12	15	9	18		0.066667	0.055556
13	18	10	20		0.055556	0.05
14	18	11	22		0.055556	0.045455
15	22	12	24		0.045455	0.041667
16	25	13	26		0.04	0.038462
17	30	16	32		0.033333	0.03125
18	30	15	30		0.033333	0.033333
19	40	20	40		0.025	0.025
20	50	23	46		0.02	0.021739
21	60	28	56		0.016667	0.017857
22	75	31	62		0.013333	0.016129
23	100	34	68		0.01	0.014706
24	130	39	78		0.007692	0.012821

Figure 4

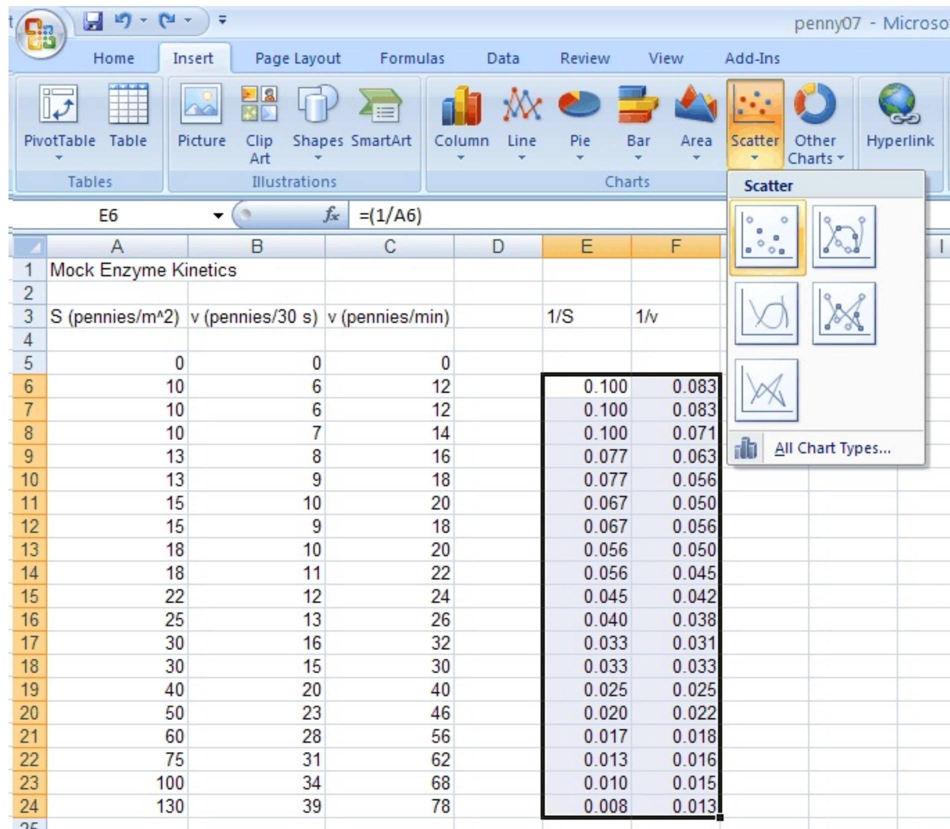
If desired, the values for 1/S and 1/v can be formatted to three decimal places to make the sheet easier to read.

- Select the data in the two columns
- Right click and choose **Format Cells**
- Click on the **Number** tab
- Under **Category**, choose **Number** and set **Decimal** places to **3**

- Click **OK**

It's time to start analyzing the data. By creating a **double-reciprocal plot** (or Lineweaver-Burk plot) the values for **K_m and V_{max}** can be determined from a **regression line** through the values for 1/S vs. 1/v. The first step is to create a **scatter plot** from the data (Fig. 5).

- Select the data under 1/S and 1/v
- Click on the **Insert** menu tab at the top of the page
- Click on **Scatter**
- Choose the "points only" plot in the upper lefthand corner
- A simple plot should be generated automatically. Select the plot and move it by clicking on the border and dragging it to a convenient location on the worksheet



	A	B	C	D	E	F
1	Mock Enzyme Kinetics					
2						
3	S (pennies/m^2)	v (pennies/30 s)	v (pennies/min)		1/S	1/v
4						
5	0	0	0			
6	10	6	12		0.100	0.083
7	10	6	12		0.100	0.083
8	10	7	14		0.100	0.071
9	13	8	16		0.077	0.063
10	13	9	18		0.077	0.056
11	15	10	20		0.067	0.050
12	15	9	18		0.067	0.056
13	18	10	20		0.056	0.050
14	18	11	22		0.056	0.045
15	22	12	24		0.045	0.042
16	25	13	26		0.040	0.038
17	30	16	32		0.033	0.031
18	30	15	30		0.033	0.033
19	40	20	40		0.025	0.025
20	50	23	46		0.020	0.022
21	60	28	56		0.017	0.018
22	75	31	62		0.013	0.016
23	100	34	68		0.010	0.015
24	130	39	78		0.008	0.013

Figure 5

Begin to format the plot by selecting it (left click on plot). At this point, a tab labeled **Chart Tools** will appear above the top menu bar (Fig. 6).

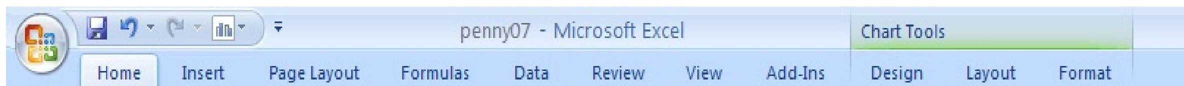


Figure 6

Select **Chart Tools** and a new menu will appear under the **Design** tab (Fig. 7).

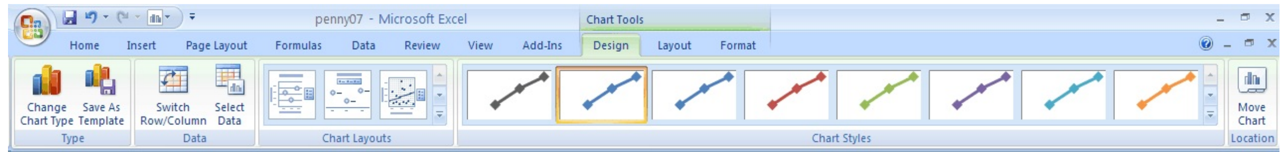


Figure 7

On the **Chart Layouts** menu left click on the arrows to get the menu of choices. Select **Layout 9** (Fig. 8)

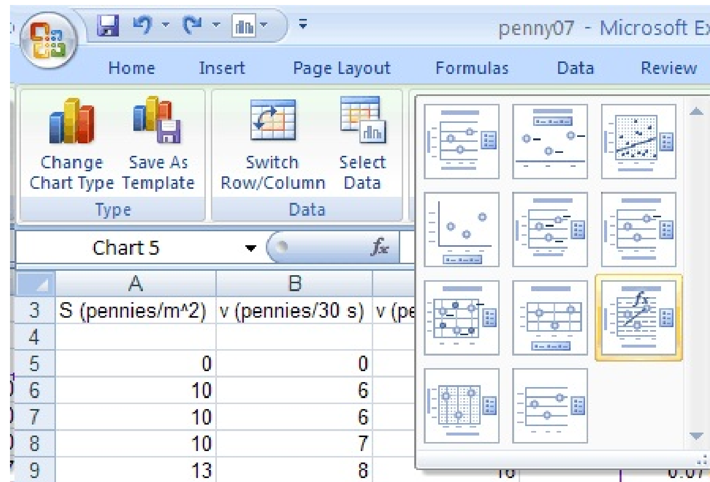


Figure 8

This will format the plot to allow the easy addition of a title and axis labels. In addition, it will draw a regression line and provide the equation of the line and its R^2 value. (Fig. 9)

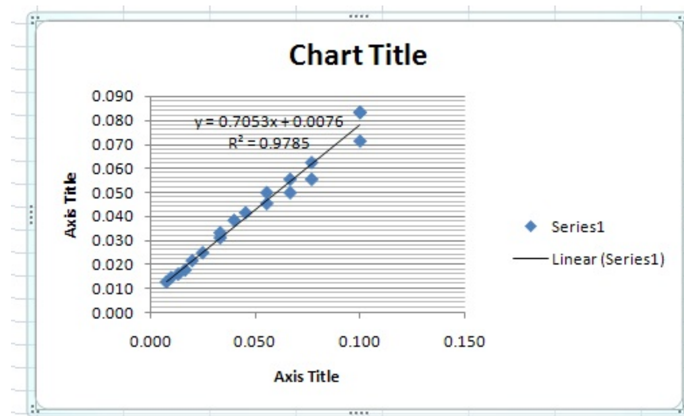


Figure 9

Enter the chart title and axis titles by left clicking on the appropriate label (click 3 times) and typing in the information. The legend on the right is unnecessary and can be deleted (Fig. 10).

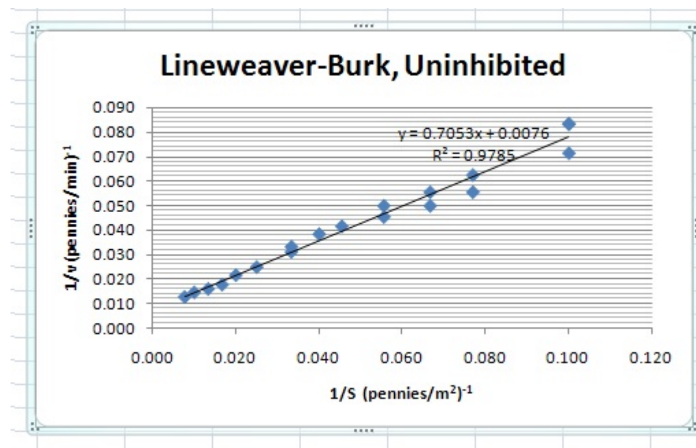


Figure 10

In the example shown above (Fig. 10), the R^2 value indicates that almost 98% of the variation in 1/v (y) is due to the variation in 1/S (x). In addition, if we take the square root of R^2 we can determine that the correlation coefficient, r, is almost 1, indicating an excellent fit between the data points and the regression line and showing that as 1/S increases, 1/v increases. The equation of the line is used to provide the K_m and V_{max} values for the enzyme. The y-intercept, 0.0076, is equal to $1/V_{max}$. Therefore, $V_{max} = 1/0.0076 = 131.579$. The slope of the regression line, 0.7053, is equal to K_m/V_{max}, so $K_m = (V_{max})(K_m/V_{max}) = (131.579)(0.7053) = 92.803$. These values can be calculated and recorded on the spreadsheet (Fig. 11).

1/Vmax =	0.0076
Vmax =	131.5789
Km/Vmax =	0.7053
Km =	92.80263

Figure 11

The values for K_m and V_{max} provide valuable information about the enzyme and can be used to plot the Michaelis-Menton Curve.

- Create a new plot, showing the relationship between **S** and **v**.
- Highlight the data in columns A and C. Select the first value in **A6** and drag down the data column. Holding the **Ctrl** key, left click and drag down the data in column C (Fig. 12)

	A	B	C
1	Mock Enzyme Kinetics		
2			
3	S (pennies/m^2)	v (pennies/30 s)	v (pennies/min)
4			
5	0	0	0
6	10	6	12
7	10	6	12
8	10	7	14
9	13	8	16
10	13	9	18
11	15	10	20
12	15	9	18
13	18	10	20
14	18	11	22
15	22	12	24
16	25	13	26
17	30	16	32
18	30	15	30
19	40	20	40
20	50	23	46
21	60	28	56
22	75	31	62
23	100	34	68
24	130	39	78

Figure 12

- As you did earlier, select the **Insert** menu from the toolbar
- Select the same type of scatter plot as before. A plot will be generated automatically.
- Move the plot to a convenient location.
- Format the plot by selecting **Layout 1** from the **Charts Layout** menu.
- Modify the title and axis labels; remove the legend (Fig. 13)

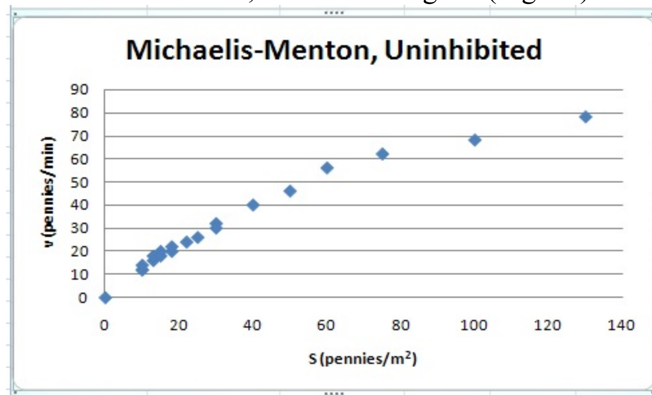


Figure 13

In order to draw an accurate line through the data points, the first step is to **calculate values for v** using the kinetic constants determined with the double-reciprocal plot.

- Using the Michaelis-Menton equation, $v = (V_{max} \bullet [S])/(K_m + [S])$, determine values of **v** for the substrate concentrations
- **Label column D "calc v"** to designate it as the calculated values
- Use the values determined for V_{max} and K_m with each value for S to calculate a corresponding value value for v. In my worksheet, the value for V_{max} is in cell J25and the value for K_m is in cell J27.
- Type the formula, **=(J$25*A5)/(J$27+A5)**, into cell **D5** (The $ in the formula is to set the row as an absolute address so that it won't change when the formula is copied to other rows.)
- **Copy** the formula in cell **D5** and paste it into the cells below (Fig. 14).

	A	B	C	D	E	F
1	Mock Enzyme Kinetics					
2						
3	S (pennies/m^2)	v (pennies/30 s)	v (pennies/min)	calc v	1/S	1/v
4						
5	0	0	0	0.0		
6	10	6	12	12.8	0.100	0.083
7	10	6	12	12.8	0.100	0.083
8	10	7	14	12.8	0.100	0.071
9	13	8	16	16.2	0.077	0.063
10	13	9	18	16.2	0.077	0.056
11	15	10	20	18.3	0.067	0.050
12	15	9	18	18.3	0.067	0.056
13	18	10	20	21.4	0.056	0.050
14	18	11	22	21.4	0.056	0.045
15	22	12	24	25.2	0.045	0.042
16	25	13	26	27.9	0.040	0.038
17	30	16	32	32.1	0.033	0.031
18	30	15	30	32.1	0.033	0.033
19	40	20	40	39.6	0.025	0.025
20	50	23	46	46.1	0.020	0.022
21	60	28	56	51.7	0.017	0.018
22	75	31	62	58.8	0.013	0.016
23	100	34	68	68.2	0.010	0.015
24	130	39	78	76.8	0.008	0.013

Figure 14

Plot the data in **column D** on the **Michaelis-Menton plot**.

- Click on the plot to select it
- **Right-hand click**
- Choose **Select Data** from the menu (Fig.15)

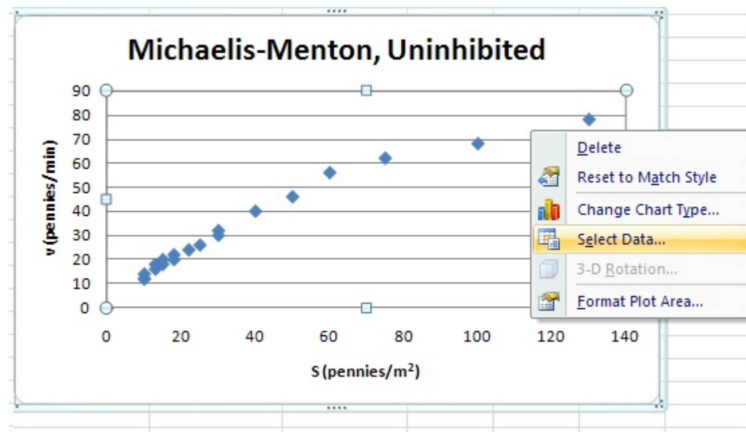


Figure 15

- On the **Select Data Source** menu (Fig. 16), left click on the **Add** button to get the **Edit Series** menu.

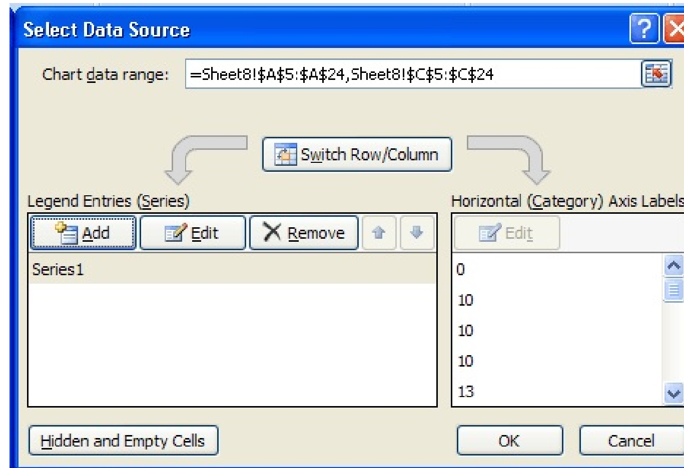


Figure 16

- Fill in the blanks on the **Edit Series** menu with the appropriate information (Fig. 17)

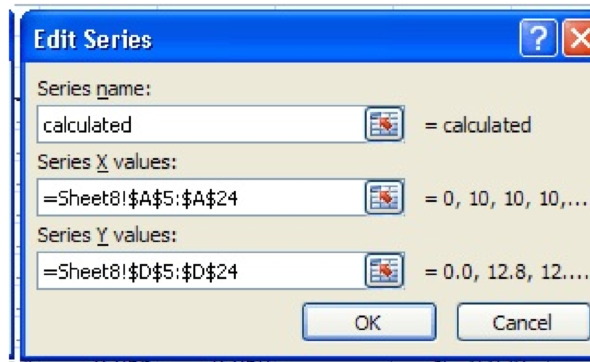


Figure 17

- In this example, the **Series name** was named "calculated." Enter the data for the X and Y values by clicking on the small boxes with the red arrows and selecting the appropriate data series from the worksheet.
- Click **OK** on the **Edit Series** and the **Select Data Source** menus and the data in the "calc v" column will be added to the plot (Fig. 18).

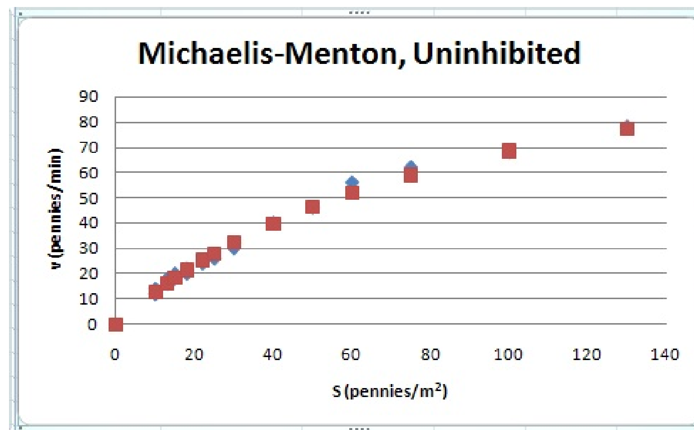


Figure 18

The data points from the calculated values need to be converted to a line.

- Move the cursor to a data point, leaving it still until a popup box appears that shows **calculated** as the series (Fig. 19)

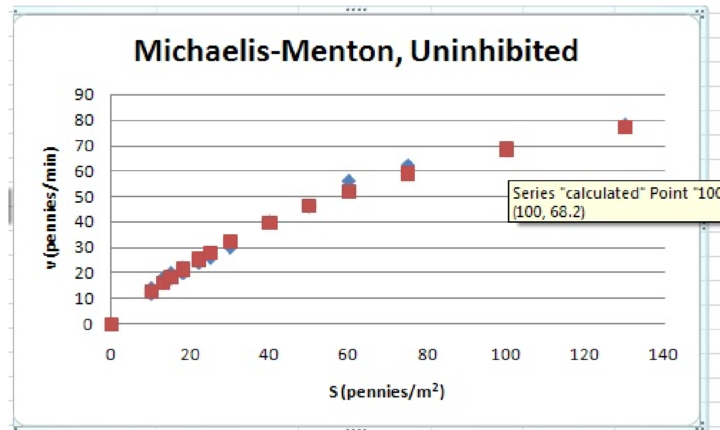


Figure 19

- **Right-click** and choose **Format Data Series** from the menu
- On the **Format Data Series** submenu, select **Marker Options** and then **None.** Now, select **Line Color** and **Solid line**.
- Click **Close** to see the modified plot on the spreadsheet (Fig. 20).

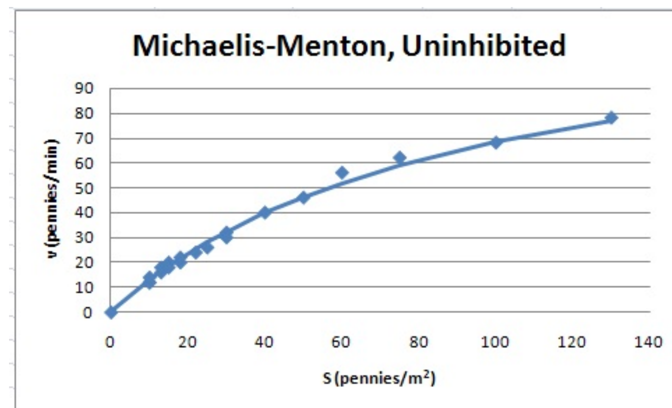


Figure 20

The Eadie-Hofstee plots can be constructed in a manner similar to constructing the Lineweaver-Burk plots. Instead of 1/S, the x-axis (and corresponding data column) will be v_o/S. The y-axis is v_o, so no further calculations are required. The y-intercept is V_{max} and the slope is -K_m.

This completes basic data analysis of the Mock Enzyme kinetic data with Microsoft® Excel. Three plots were produced from each assay, one representing the Michaelis-Menton equation and the other two representing linearized forms of that equation, specifically a double-reciprocal plot called a Lineweaver-Burk plot and an Eadie-Hofstee plot. Values for K_m and V_{max} should be determined from each plot and compared. Although the values should be similar between plots of the same data, they may not be. Discuss why this may

be the case and explain which plot provides more accurate values for K_m and V_{max}. (You may want to look at curves on the Michaelis-Menton plot generated from both sets of values.)

A more detailed analysis of your data can be performed by selecting **Regression** from the **Data Analysis** menu (**Data > Data Analysis > Regression**). The **Regression** menu (Figure 21) is straight-forward and the choices allow for examination of several aspects of the data. In this example the confidence interval is left at the default value of 95% and all of the analytical options have been selected. Using the data for the Lineweaver-Burke plot above, several tables and plots are generated (Figures 22and 23).

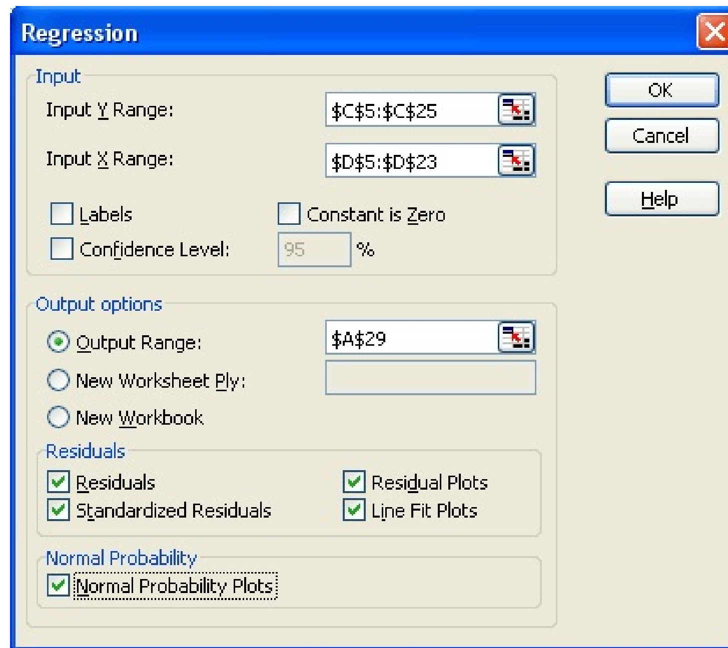


Figure 21

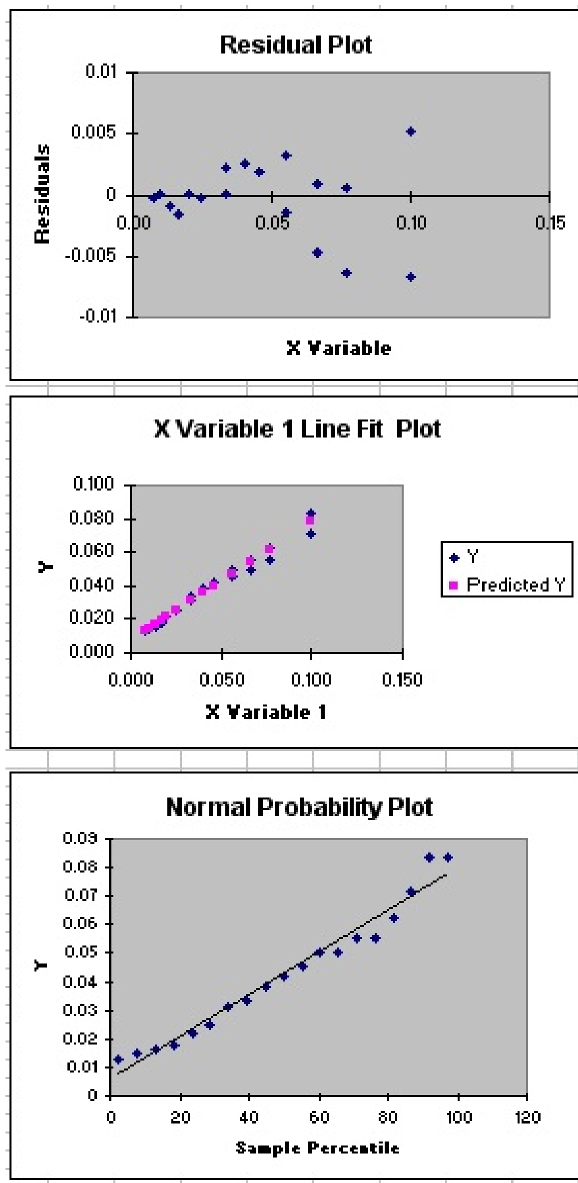


Figure 22

Appendix 4. Using Microsoft® Excel to Analyze Data from the Disk Diffusion Assay

Entering and Formatting Data

Open Excel. Set up the spreadsheet page (Sheet 1) so that anyone who reads it will understand the page (Figure 1).

- Type a **title** in the cell in the upper lefthand corner, cell A1
- **Label column A** as the data from the 0.5 McFarland culture with gentamicin (GM10) in cell A3
- **Label column B** as the data from the other McFarland culture with gentamicin. (In this example, the 1 McFarland culture.)
- Repeat with the trimethoprim/sulfamethoxazole (SXT) data.
- **Enter the appropriate data** (diameter of zones of inhibition in mm) in each column (Fig. 1).

	A	B	C	D
1	Disk Diffusion Assay			
2				
3	GM10, 0.5	GM10, 1	SXT, 0.5	SXT, 1
4	24	24	30	29
5	25	26	31	31
6	24	25	30	32
7	25	24	31	31
8	25	25	31	31
9	24	26	30	34
10	24	25	31	30
11	25	26	30	32
12	24	25	30	32
13	24	25	30	32
14				

Figure 1

Calculating the Mean and Standard Deviation (should this be desired)

You can **calculate the mean and standard deviation** for each column using the functions "**=average( )**" and "**=stdev( )**", respectively (Fig 2).

3	GM10, 0.5	GM10, 1
4	24	24
5	25	26
6	24	25
7	25	24
8	25	25
9	24	26
10	24	25
11	25	26
12	24	25
13	24	25
14		
15	=average(A4:A13)	

Figure 2

- The **data for each column goes inside the parentheses** and may be entered by left-clicking and dragging the mouse over the values after typing the open parenthesis.
- Finish by typing the **close parenthesis** and hitting the "Enter" button.
- **Copy the functions** (left-click and drag to select, right click and "Copy") and paste them under the other columns (left-click and drag to select, right click and "Paste") (Fig 3).

	A	B	C	D
1	Disk Diffusion Assay			
2				
3	GM10, 0.5	GM10, 1	SXT, 0.5	SXT, 1
4	24	24	30	29
5	25	26	31	31
6	24	25	30	32
7	25	24	31	31
8	25	25	31	31
9	24	26	30	34
10	24	25	31	30
11	25	26	30	32
12	24	25	30	32
13	24	25	30	32
14				
15	24.4	25.1	30.4	31.4
16	0.516398	0.737865	0.516398	1.349897

Figure 3

The *t*-test

The easiest way to perform **statistical analyses** with a spreadsheet is to **use the built-in functions**. With Excel, they can be found by **clicking on "Tools"** on the toolbar and **selecting "Data Analysis"** (Fig. 4).

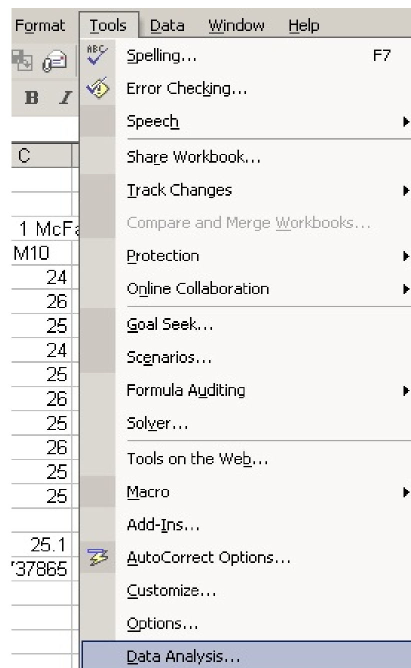


Figure 4

If the "**Data Analysis**" selection is not listed it means that the functions haven't been installed.

To install them, you must have access to the **installation program**, either on CD or through a network.

- Select "**Add-Ins**" under the "**Tools**" menu (Fig. 4).
- When the "Add-Ins" menu comes up, choose "**Analysis ToolPak**" (Fig. 5) and click on "OK."

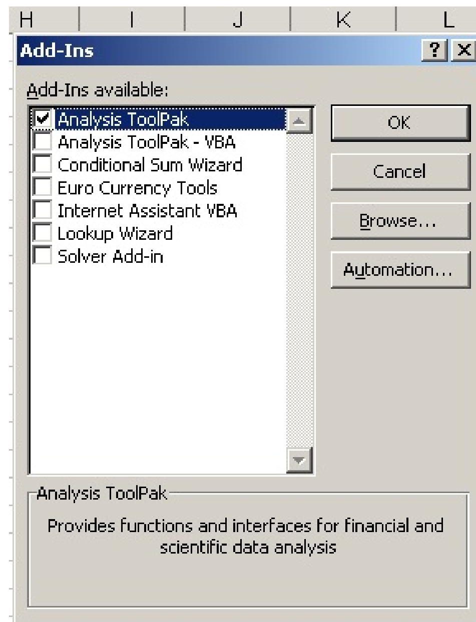


Figure 5

When you select "Data Analysis" on the "Tools" menu (Fig. 5), the "Data Analysis" menu pops up (Fig. 6).

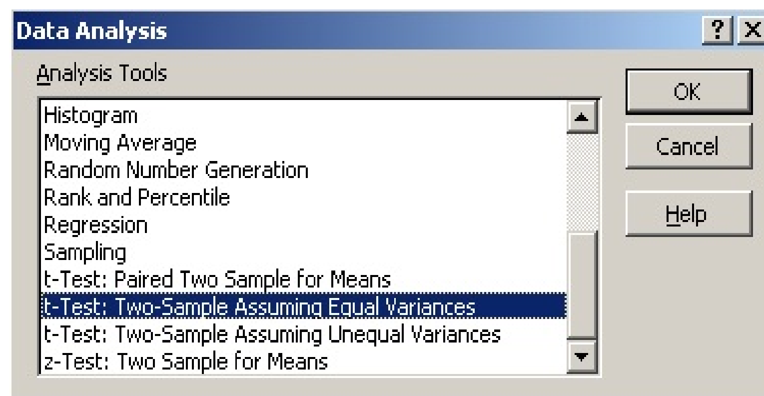


Figure 6

- Scroll down and **select "t-test: Two-Sample Assuming Equal Variances."**
- After you **click "OK,"** the "t-test: Two-Sample Assuming Equal Variances" menu pops up (Fig. 7) and you need to **select your data**.

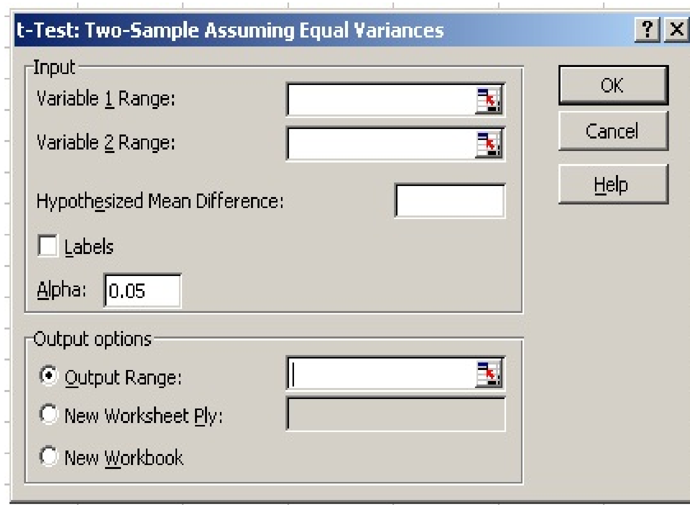


Figure 7

- **Click on the red arrow** on the right side of the box next to the "**Variable 1 Range:**" label.
- A menu pops up for data input (Fig. 8).

Figure 8

- **Enter data** by dragging the mouse over the values in the appropriate column, for example Column A, cells 3-13. **Hit "Enter" when done**. This will input the values for the zones of inhibition with gentamicin (GM10) on the plates inoculated from the 0.5 McFarland culture, as well as the column label.
- **Repeat this process**, clicking on the arrow for "**Variable 2 Range:**" and entering the other gentamicin data in Column B.
- Leave the "**Hypothesized Mean Difference**" selection blank, check the "**Labels**" box, and leave "**Alpha**" at 0.05.
- **Select a cell** on the spreadsheet, for example F3, where you want the **results of the *t*-test** to be placed.
- The completed "**t-test: Two-Sample Assuming Equal Variances**" menu should look

similar to Figure 9.

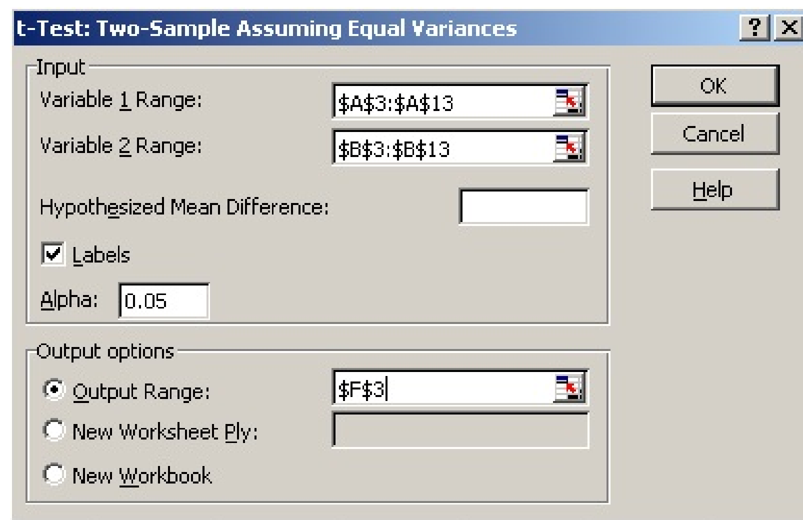


Figure 9

- **Click on "OK"** and the table of information for the *t*-test should appear, beginning in cell F3.
- Move the cursor to the right hand border of the Column F label so that the **column resize cursor** appears (Fig. 10).

column resize cursor

F	G	H
t-Test: Two-Sample Assuming Equal Variances		
	GM10, 0.5	GM10, 1
Mean	24.4	25.1
Variance	0.266667	0.544444
Observatio	10	10
Pooled Va	0.405556	
Hypothesi:	0	
df	18	
t Stat	-2.45786	
P(T<=t) on	0.012172	
t Critical oı	1.734063	
P(T<=t) tw	0.024345	
t Critical tv	2.100924	

Figure 10

- Drag the column to the right until the labels can be read (Fig. 11).
- **Compare the calculated *t* value (t Stat) with the critical value for *t*** (t Critical) for alpha = 0.05 and with 18 degrees of freedom (df). We might expect that with a higher concentration of cells the zone of inhibition would be equal to or smaller than the zone produced with the 0.5 McFarland culture, so a one-tailed *t*-test would be appropriate.

t-Test: Two-Sample Assuming Equal Variances		
	GM10, 0.5	*GM10, 1*
Mean	24.4	25.1
Variance	0.266667	0.544444
Observations	10	10
Pooled Variance	0.405556	
Hypothesized Mean Difference	0	
df	18	
t Stat	-2.45786	
P(T<=t) one-tail	0.012172	
t Critical one-tail	1.734063	
P(T<=t) two-tail	0.024345	
t Critical two-tail	2.100924	

Figure 11

Therefore, if the calculated *t* value (-2.45786, in this example) is less than the negative critical value for *t* (-1.734073), we reject the null hypothesis that there is no difference between the zones of inhibition from the 0.5 and 1 McFarland cultures.

ANOVA

To **set up a spreadsheet for ANOVA**, list the measurements of zones of inhibition made by each student according to antibiotic and McFarland culture (Fig. 12). Calculate means and standard deviations, if desired.

	A	B	C	D
1	Disk Diffusion			
2				
3	SXT, 0.5			
4	A	B	C	D
5	29	29	30	27
6	32	29	31	31
7	29	29	30	27
8	30	32	31	31
9	30	31	31	30
10	30	32	30	31
11	30	31	31	30
12	30	30	30	30
13	29	29	30	31
14	29	30	30	30
15				
16	29.8	30.2	30.4	29.8
17	0.918937	1.229273	0.516398	1.549193

Figure 12

For **ANOVA**, use the "**Data Analysis**" menu under the "**Tools**" menu.

- Click on "Tools" on the toolbar
- Select "Data Analysis" (Fig. 4)

- Choose “**ANOVA: Single Factor**” (Fig. 13).

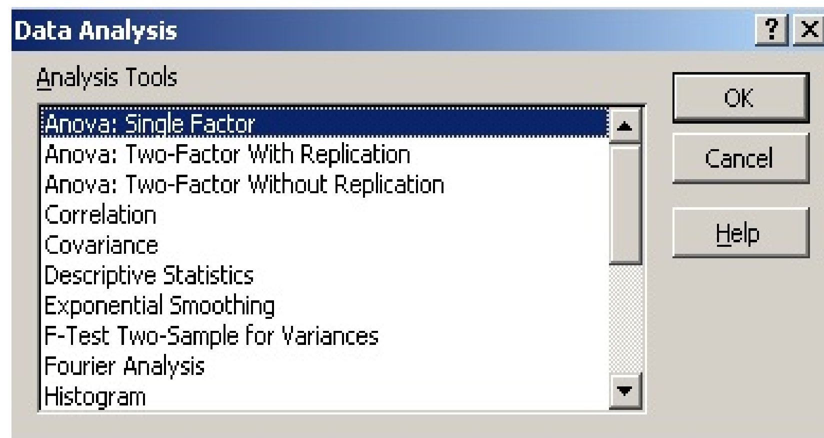


Figure 13

- For “**Input Range**:” select the **data in all four columns**, including the headings in Row 4.
- Make sure that “**Columns**” is selected in “**Grouped By:**.”
- Select “**Labels in First Row**.”
- “**Alpha:**” should be 0.05.
- For “**Output Range**” select a cell near the data. (In this example, F3.)
- When complete, the “ANOVA: Single Factor” menu should look like Figure 14. Click “OK” and a table will be generated, starting with Cell F3 (Fig. 15).
- **Adjust column widths** so that the table looks like Figure 15.

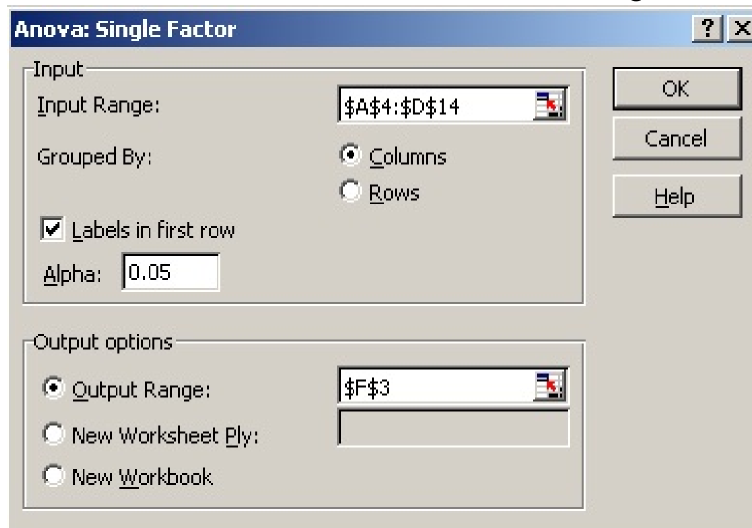


Figure 14

- **Compare the calculated *F* value** (0.716814, in this example) with the **critical *F* value**

(*F crit*)(2.866265). Since the *F* value is less than the critical *F* value, we cannot reject the null hypothesis that there is no difference between the measurements of zones of inhibition made by each of the four group members.

- Nowadays, the more appropriate way to evaluate the results of a statistical test is to observe the *P*-value. In this case the value, 0.548441, is greater than the alpha value, 0.05, so we conclude that the null hypothesis cannot be rejected and there is no difference between the measurements of zones of inhibition made by each of the four group members.

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
A	10	298	29.8	0.844444		
B	10	302	30.2	1.511111		
C	10	304	30.4	0.266667		
D	10	298	29.8	2.4		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	2.7	3	0.9	0.716814	0.548441	2.866265
Within Groups	45.2	36	1.255556			
Total	47.9	39				

Figure 15

Appendix 5. Looking for Relationships

One of most common types of investigation we do is to look for relationships between variables. Variables may be nominal (categorical), for example looking at the effect of an antibiotic on gram positive vs. gram negative bacteria, or numerical, for example the effect of a nutritional supplement on growth. Nominal data have observations for at least two groups in mutually exclusive categories that are compared (*e.g.,* male/female, heterotrophic/autotrophic, aerobe/facultative anaerobe/aerotolerant anaerobe/microaerophile/anaerobe, motile/nonmotile, etc.). The most common

Q: What is *nominal* data?
A: Nominal data is data that can be assigned to a category, rather than data that can be described by a numerical value.

test used to determine if there is an association (or independence) between at least two groups with observations that can be assigned to two or more categories is the *chi-square test of association* (sometimes called a test of independence). (The chi-square test is also used as a test of goodness-of-fit where observations are compared to theoretical values, like phenotypes in genetic crosses.) So, the chi-square test compares counts of categorical data between two or more variables and can only be applied to actual numbers, not percentages, statistics, or data that has been transformed in any manner. The null hypothesis is that the categories for each of the variables are independent. In other words, there is no association between observations in a category for one variable and observations in that same category for another variable. An alternative hypothesis is that the distribution of categorical variables differ from each other. For example, if we were looking at the type of vehicle driven by microbiologists and mathematicians, the null hypothesis could be that the proportion of microbiologists who drive trucks is the same as the proportion of mathematicians who drive trucks (the ratios are the same). The alternative hypothesis is that the proportions of trucks driven by microbiologists and mathematicians are not the same. That is, there is some association between professional backgrounds and the type of vehicle driven.

Q: What is the chi square test of association (or independence)?
A: The chi square test is a non-parametric test used to determine if there is an association between categorical variables.

The chi-square test is fairly simple to perform and can be set up very easily in a spreadsheet. For two groups, each with two categories, we set up a *2x2 contingency table* (Table 1). *Contingency* means *possibility*, so a contingency table is just a table of possibilities.

Table1. Sample contingency table.

	Category 1	**Category 2**	**Total**
Group 1	a	b	**a + b**
Group 2	c	d	**c + d**
Total	**a + c**	**b + d**	**a + b + c + d**

Q: What is a contingency table?
A: A contingency table is a table that presents all the possible combinations of two or more variables.

To calculate the chi-square value, the first step is to calculate the expected values for each category of each group. This is done for each cell by multiplying the sum of the corresponding row by the sum of the corresponding column and dividing that product by the total of all the cells (Table 2). Once the expected values are calculated, the chi-square value is calculated for each cell by subtracting the expected value from the observed value, squaring that value, and dividing it by the expected value. The chi-square value is calculated for each cell and totaled to give the chi-square value for all the data (Eqn.1). That chi-square value is compared to a critical chi-square value determined for the appropriate degrees of freedom and α level of significance. The degrees of freedom would be (number of data columns -1) x (number of rows -1).

Table 2. Calculation of expected values.

	Category 1	**Category 2**
Group 1	(a+b)*(a+c)/(a+b+c+d)	(a+b)*(b+d)/(a+b+c+d)
Group 2	(c+d)*(a+c)/(a+b+c+d)	(c+d)*(b+d)/(a+b+c+d)

$$\chi^2 = \sum \frac{(observed - expected)^2}{expected} \quad (1)$$

For a 2x2 contingency table this becomes (2-1) x (2-1) = 1. We usually assign α = 0.05. From a table of chi-square values (Table 3), we find that the critical value for 1 degree of freedom and α = 0.05 is 3.841. If the calculated value for chi-square is 3.841 or less we accept the null hypothesis and say that there is no difference between the groups. Alternatively, if using Microsoft® Excel, we can use the CHIDIST function to calculate a P value.

A limitation of the chi-square test is that it should not be used if an expected value is less than 5. When that occurs, it is better to use the Yate's correction (subtract 0.5 from the difference between observed and expected values) or Fisher's exact test.

Table 3. Chi square values

	α			
df	**0.1**	**0.05**	**0.01**	**0.001**
1	2.706	3.841	6.635	10.827
2	4.605	5.991	9.210	13.815
3	6.251	7.815	11.345	16.268
4	7.779	9.488	13.277	18.465
5	9.236	11.070	15.086	20.517

Q: What does it mean to say that two variables are not associated?
A: Independence suggests that the variables are from different populations and that their lack of association did not occur by chance.

Numerical variables may be discrete or continuous, that is they are not nominal, and associations can be determined with correlation analysis or regression analysis. (Remember, discrete

numerical values are whole numbers, for example the number of children in a household, while continuous values can have any value within a range, for example the height of those children.) Correlation analysis is used to determine if there is an association between two variables and, if so, how strong that association is. Regression analysis does the same but also assumes a cause and effect relationship. Correlations may be positive (as one variable increases, so does the other) or negative (as one variable increases, the other decreases). In any examination of the relationship between two (or more) factors, there are three possible interactions:

1. changing one variable produces a consistent change in the other
2. changing one variable has no consistent effect on the other variable, so that any changes in the second variable appear to be a coincidence
3. both variables change in response to third variable.

The first interaction describes a *causal* relationship where a change in one variable (or factor) has a predictable effect on the second variable, hence the term "cause and effect." The strength of this association can be calculated using Pearson's Correlation Coefficient or by performing a linear regression.

The second interaction is really no interaction. In other words, there is no association between the variables and any changes that occur in one variable when the other is changed only occurs by chance.

The third interaction is fairly common and often misinterpreted as a causal interaction. (**Correlation does not mean causation!**) A classic example is that as the number of churches in a town increases, so does the number of bars. An interesting observation but can we demonstrate a direct cause and effect relationship? Probably not. Instead, it seems much more likely that both variables are influenced by the size of the town. That is, more people = more churches = more bars. (How might we attempt to determine if there really is a causal relationship between the number of churches and the number of bars?) So, an important point to remember and one that is easily overlooked is that **correlation does not mean causation!**

In a correlation analysis, samples are selected at random, variables are assumed to be normally distributed, and a line is **not** drawn through the plots. The most commonly used correlation analysis is calculation of the Pearson correlation coefficient; the symbol for the parameter is ρ and for the statistic is r. We refer to r as the correlation coefficient. The Pearson correlation coefficient can be used when:

a. the sample is a random sample from a population
b. both variables are normally distributed
c. measurements of both variables are on an interval or ratio scale (continuous)
d. the relationship between the two variables is linear.

The value of r varies from -1 (perfect negative correlation) to +1 (perfect positive correlation). The null hypothesis is of no association ($r = 0$). The correlation coefficient is not a measure of significance. Significance can be calculated with the *t*-test. For the *t*-test, degrees of freedom is n-2 and the equation is:

$$t = r\sqrt{\frac{n-2}{1-r^2}} \qquad \textbf{(2)}$$

The critical value is obtained from the table of the *t*-distribution using the appropriate values for degrees of freedom and α. In some cases, a table of critical values for r may be available and can be used directly. In that case, the correlation is significant if the calculated value of r is equal to or greater than the tabular value (null hypothesis is rejected).

Q:	What symbol represents the correlation coefficient and what is its range?
A:	The correlation coefficient is symbolized by *r* and it ranges from -1 (perfect negative correlation) through 0 (no correlation) to +1 (perfect positive correlation).

If at least one of the variables is not normally distributed a non-parametric test should be used. The most common non-parametric test used is Spearman's rank correlation test. In this test, each variable is ranked from smallest to largest and, for each data pair, the difference in rank is calculated (x-rank - y-rank). The differences are summed and squared.

Simple linear regression is similar to correlation analysis but a cause and effect relationship is assumed. This cause and effect relationship can be tested with an experimental design and, if demonstrated, allow us to predict a value of *y* (dependent variable) that corresponds to a given value of *x* (independent variable). The relationship between the variables is described as a straight line with the equation $y = a + bx$, where *a* is the y-intercept (point on the y-axis where the line intersects) and *b* is the slope of the line. Determining the values of *a* and *b* allow us to plot the regression line and predict values of one variable from the other.

Q:	What is the equation for a straight line and what do the symbols mean?
A:	The equation for a straight line is $y = a + bx$. The symbol *y* is the variable dependent of the value of *x*; *a* represents the point on the y-axis where the line intercepts ($x = 0$) and *b* is the slope of the regression line.

The assumptions for performing simple linear regression are:

1. Values of the independent variable are fixed. That is, they are chosen by the investigator, are therefore not random, and have no associated variance.
2. For any value of the independent variable, there is a normally distributed population of values for the independent variable, described as $y = a + bx$.
3. For any value of *x* there is a value of *y* such that $y = a + bx + e$ where *e* is the residual or the amount that the observed *y* differs from the mean value of *y*. Residuals are normally distributed.
4. The variances of *y* for all values of *x* are equal.
5. Observations are independent. That is, each individual in the sample is only measured once.

The coefficient of determination (r^2) is a measure of the proportion of variance in the dependent variable (y) that is due to its dependence on x. (This is also called the strength of association.) Remember, one of the assumptions for simple linear regression is that the values of the independent variable are selected by the investigator and therefore have no variance. As the value of y changes in response to changing values of x, calculation of r^2 provides us with an estimate of how much of the variation in y is due to x (as opposed to other sources of variation like measurement error). The higher the value of r^2 (maximum of 1, minimum of 0), the stronger the dependence (association) of the independent variable on the dependent variable. The coefficient of determination can be calculated for correlation analysis by squaring the regression coefficient, r, but remember that correlation analysis does not show a cause and effect relationship so in this case r^2 is not a measure of variation in y explained or dependent on x, but instead is a measure of the strength of the association between the variance in y with the variance in x.

Q: What is the difference between correlation and linear regression?
A: Correlation is a measure of association between two variables without assuming that either one is dependent on the other. Linear regressions examine a cause and effect relationship and is determined experimentally.

Excel

1. There are two ways to perform a linear regression in Excel. The simpler method is to use the XY (scatter) plot of the data points.
   a. After creating the scatter plot, right-click on the data points and get a pop-up menu.
   b. Select "Add Trendline" and the "Add Trendline" menu pops up (Figure 1).
   c. Select "Linear" on the "Type" page. On the "Options" page select "Display equation on chart" and "Display R-squared value on chart" before selecting the "OK" button.

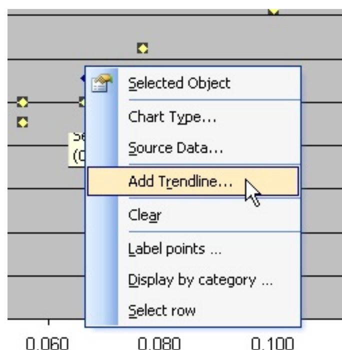


Figure 1. Pop-up menu to add a trendline to an XY plot.

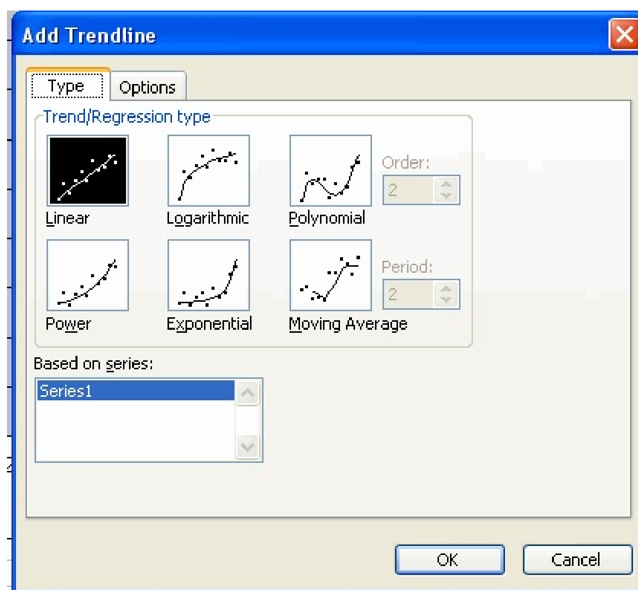


Figure 2. Menu for selecting a linear trendline.

   d. The chart should now display a straight (regression) line through

the data points and provide the equation of the line and the r^2 value on the chart. In the example shown (Figure 3), the slope of the line is 0.7053, the y-intercept is 0.0076, and the r^2 value is 0.9785, indicating that the variance of y is almost completely due to the changes in the value of x.

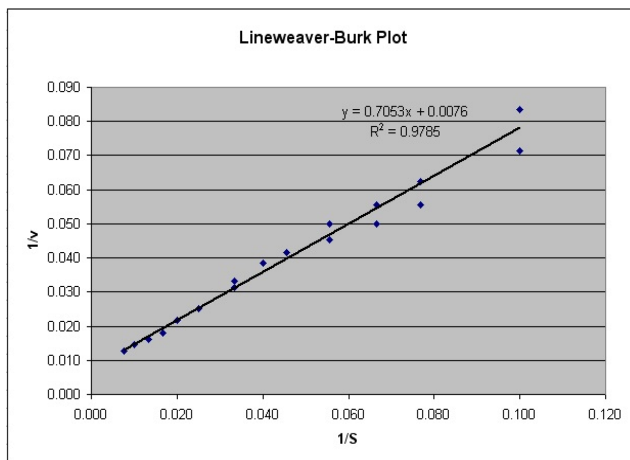


Figure 3. Sample regression line.

This method is convenient since a plot of the data will be constructed anyway, but it lacks quite a bit of information. To obtain that information, use the statistics function "Regression" located in "Data Analysis" under"Tools" in the menu bar. (If "Data Analysis" is not listed, it needs to be installed using the "Add-Ins" selection under "Tools.")

2. Once "Regression" has been selected, the "Regression" menu box pops up. Enter the data ranges for y and x in the appropriate locations. The "Confidence Level" should be 95% (the default value). Select where the table of calculated values should be written (*e.g.,* "New Worksheet Ply"). Select "OK" and the data will appear under a heading of "SUMMARY OUTPUT", in this case in a new worksheet (Figure 4).
3. The "SUMMARY OUTPUT" is divided into three sections. Under "Regression Statistics" we find the value for the correlation coefficient, r, as "Multiple R." The value for the coefficient of determination, r^2, is listed as "R Square."
4. The slope of the regression line is listed in the bottom table as "X Variable 1." The y-intercept is listed right above that as "Intercept."

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.98918788							
R Square	0.97849266							
Adjusted R Square	0.97722752							
Standard Error	0.00336577							
Observations	19							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	0.008761704	0.008762	773.4277	1.29283E-15			
Residual	17	0.000192583	1.13E-05					
Total	18	0.008954287						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.007627	0.001476854	5.164359	7.78E-05	0.004511114	0.0107429	0.00451111	0.010742891
X Variable 1	0.70534216	0.025362379	27.81057	1.29E-15	0.651832217	0.7588521	0.65183222	0.758852102

Figure 4. Summary output table from regression analysis.