Chapter 3

The Scientific Method:
An Introduction Using Reaction Time

Bob Kosinski and John Cummings

Biology Program, Clemson University
Clemson, SC 29634-1902
(864) 656-3830
Fax: (864) 656-3839
rjksn@clemson.edu
cumminj@clemson.edu

Bob Kosinski received his BS in Biology from Seton Hall University and his Ph.D. in Ecology from Rutgers University. He is a Professor of Biology at Clemson University, where he teaches introductory biology. His interests include investigative laboratories in introductory biology, and the role of computer simulations and physiological interfacing in introductory biology.

John Cummings is the laboratory coordinator for introductory biology courses at Clemson University. He received his BS and MS degrees from Bowling Green State University. His professional interests include teaching in HELL (high-enrollment laboratory learning), the training of teaching assistants, and barn owl biology.

Contents

Introduction

The purpose of this exercise is to introduce the scientific method, experimental design, and an elementary statistical analysis to freshman biology students in a one-period exercise. Using a computer program that measures reaction time, the students are asked to propose, execute, and analyze an original experiment. The fact that the variable being tested is reaction time introduces an element of competition that heightens student interest.

The only equipment the exercise requires is a Windows-compatible computer or Macintosh computer and Kosinski's *Reaction Time* program. Windows and Macintosh versions of this program have been published by Freeman Publishing Company (Kosinski, 1998). The program has two parts: a reaction time portion, and a statistics portion.

In the reaction time portion, the students must press the spacebar as soon as they see or hear a stimulus from the computer. There are five different types of stimuli, ranging from the simple (for example an X appears in the middle of the screen or a tone sounds) to the complex (a letter will appear in the middle of the screen, but the user should only press the spacebar if it is a certain letter).

In the statistics portion, the data are automatically ported to a program that uses an easily-explained, robust test (the chi-square median test) to determine if the treatments are significantly different. If the program is used in this way, the students never have to enter data or do any computations. Alternatively, the students could type data in manually and receive the same analysis. This component gives statistically naive students an introduction to hypothesis acceptance and rejection. While the exercise's treatment of statistics is elementary, graduate student teaching assistants have told us that it helped them understand their high-level statistics courses for the first time.

The experiments the students could propose are infinitely varied (samples are listed in Appendix B), but can be categorized into several types:

1. Testing one group of people against another (e.g., men vs. women) on the same reaction time test;
2. Comparing the reaction times of the same people on two different reaction time tests;
3. Comparing reaction times on the same test before and after some treatment such as exercise or drinking caffeine.

The student text helps the students define their variables and treatments and avoid problems such as inadequate replication and bias. Finally, at Clemson we evaluate the student on a report. A literature review on aspects of reaction time that commonly appear in the reports is provided in Appendix C. At this writing, the literature review is also available on the World Wide Web at http://biology.clemson.edu/bpc/bp/Lab/110/reaction.htm.

We have used this exercise for several years in both our majors and our nonmajors' freshman general biology course at Clemson University. The exercises require about two hours to complete; the most time-consuming part of the laboratory is the explanation of the scientific method and statistics. Student reaction to the exercises is good, both because of the inherently interesting nature of reaction time and because the exercise acts as an "icebreaker" by introducing students to teamwork and each other. Students have no difficulty using the software or devising investigations, although at times they want us to tell them the "right" investigation to do.

The greatest difficulties are inescapable because they stem from the abstract nature of the material. It is difficult for students to understand hypothesis testing and especially difficult for them to understand the principles of statistics. We've found that these are lessons that must be reinforced throughout a laboratory course. Therefore, we begin with the exercise printed here, but follow up with hypothesis-testing and reports on several different topics.

Materials

The only materials required are Windows-compatible computers or Macintosh computers and the *Reaction Time* software. One computer per four to six students works well. The group at one computer cannot be too small or it will not be able to produce enough observations or have a good discussion.

If use of the audio stimuli in the program is desired, each computer will also need a pair of earphones or speakers connected to the audio output of the computer (not the audio output of the computer's CD player).

Windows95 and Macintosh versions of the Kosinski reaction time program have been published by the W. H. Freeman Publishing Company in association with the laboratory manual *Biology in the Laboratory*, 3rd Edition (Helms et al., 1998). This "BioBytes" CD (Kosinski, 1998) comes with a student workbook for BioBytes 3.1 and contains *Reaction Time* and five other programs by Kosinski. These programs are simulations of cardiopulmonary physiology, human reproductive physiology, plant competition and plant physiology, the foraging ecology of a small shark, and population genetics. This "BioBytes" CD and a student workbook can be purchased from:

W. H. Freeman and Company
41 Madison Avenue
New York, NY 10010
1-800-877-5351

Also, Freeman supports a BioBytes website at
http://www.whfreeman.com/biolab/helms/biobytes.html.

Notes to the Instructor

There are no safety concerns in this laboratory.

As mentioned above, the major difficulties with this laboratory are conceptual; students have a difficult time with hypothesis testing and statistics.

Aside from this, an infrequent problem is failure of the statistical test to give valid results because the students didn't follow directions. This is frustrating because the students have just gone through all their reaction time experiments, and then may have to start all over again. The most common student lapses are:

Failing to Enter Data Points under the Right Treatment

In the exercise presented here, before each group of observations, the students must specify to which treatment (usually either 1 or 2) these data belong. If they make a mistake and indicate that treatment 2 observations belong to treatment 1 before the reaction times are recorded, they can still save the situation by answering "No" to the question, "Add these observations to Treatment 1?" after the reaction times are finished. This will cause this latest group of reaction times to be discarded. If they have already added the observations to the incorrect treatment, they must end the experiment and start recording reaction times again.

Not Having Individuals in the Same Order in Both Treatments When the Experiment Uses a Paired Test

A paired test compares the "Treatment 1" and "Treatment 2" reaction times of the *same individuals*. Therefore, the program assumes that if the first reaction time in Treatment 1 belongs to Student A, the first reaction time in Treatment 2 belongs to Student A also. The order of students (and the number of observations per student) must be the same in Treatment 1 and Treatment 2. This does not mean that every student must do Treatment 1 first and then Treatment 2, but the order of students must be the same in both treatments.

If the students make this mistake, if they have copied down their individual observations, they can enter them manually into the statistical program by selecting "Perform Statistical Analysis" at the Reaction Time title screen. Manual entry brings up another problem, however.

Removing Data Points Entered through Manual Data Entry

If the students enter data points manually (which will not be necessary if all the directions in this exercise are followed), they will run afoul of a bug in the published program if they try to *remove* an already-entered data point. Using directions on the screen, they can *change* the data point, but they cannot remove it. If they need to remove a data point from a treatment, they should start data entry over again.

Audio Stimuli

In a noisy laboratory, audio stimuli either require speakers hooked to the computer or a pair of earphones hooked to the sound output of the computer (not to the computer's CD player). Several speakers sounding in the same room can be confusing, so earphones are recommended. We have found that with cheap earphones, the audio stimuli are very faint, so use good-quality earphones.

Student Outline

The scientific approach is a powerful method of understanding the natural world because it is founded on our observations of how the world works. However, not just any observations will do. To be successful, the observations must be systematic and objective. The method of doing these observations is sometimes broken down into a series of steps.

The Steps of the Scientific Method

Let's say we wanted to find out whether a combination of anti-AIDS drugs X and Y is more effective in treating AIDS than either of the drugs given separately. We might proceed as follows:

1. **Use our general and perhaps non-systematic observations to devise a question about the observed system.**
    For example, "Drug X and drug Y both delay the development of full-blown AIDS in individuals infected with HIV, the AIDS virus. It seems sensible that a treatment that combines these two drugs would be even more effective. On the other hand, would the two drugs interfere with each other in some way?"
2. **Define the experiment's independent, dependent, and standardized variables.**
    The *independent variable* is the factor that is being manipulated in the current experiment. In this case, the independent variable is type of drug treatment. In the experiments you will be performing in this lab, you will usually deal with only one independent variable at a time.
    The *dependent variable* is the aspect of the system that is showing response to the manipulations of the independent variable. Here, the dependent variable could be any of the many aspects of patient condition that define the difference between full-blown AIDS and relative health.
    The *standardized variables* are all the variables that are held constant between the treatments. In this case, the way the drugs or drug mixtures are administered, the frequency at which the patients are checked, and the average ages and health of the people assigned to the treatments, are all standardized variables.
3. **Define the experiment's treatments.**
    A treatment is a group that is subjected to the same levels of the independent variable. In this case, the group that gets drug X alone is one treatment, the group that gets drug Y is another treatment, and the group that gets both drug X and drug Y is the third treatment.
4. **Devise a research hypothesis and use it to make a prediction that is testable by experiment.**
    A research hypothesis is an assertion about the way the studied system works. For example, "Drug X and Drug Y will have different effectiveness when given together than when given separately." Because we think the drugs might be more effective together, we may predict, "Treating healthy but HIV-positive individuals with both drug X and drug Y will postpone the development of full-blown AIDS significantly longer than treating a similar group with drug X alone or drug Y alone." If this prediction does not turn out to be true, this does not invalidate the experiment.

    Hypotheses don't have to be correct, but they do have to make predictions that are precise and testable by experiment.

For statistical reasons, to be covered later, you should also devise a "null hypothesis," or statement of what would happen if there is *no* effect of the treatments. In this case, the null hypothesis would be that Drug X and Drug Y will be have the same effectiveness when given alone as when given together. This null hypothesis would predict that the onset of full-blown AIDS will occur at the same time in all three treatments.

5. **Devise an experiment to test the research hypothesis.**

Take a large group of healthy but HIV-positive individuals. Randomly assign them to groups that will get drug X alone, drug Y alone, and both drugs X and Y. Make sure that you give all the treatments a fair chance against one another by seeing that they all have subjects with a similar range of ages, health, previous treatment histories, and so forth. If you don't do this, and one treatment ends up with most of the younger, healthier patients and the other with most of the older, sicker ones, you have introduced "bias" into the experiment. It will be impossible to say whether the results of the experiment were due to the drugs or to the biased selection of people who went into each treatment.

6. **Perform the experiment and collect the data.**

Make sure that all the health workers who will be giving the drugs understand the drug dosages, that they are all using the same definition of "onset of full-blown AIDS," etc. Pre-plan procedures for dealing with inevitable problems like patients who miss drug treatments or drop out of the program.

7. **Analyze the data and determine whether the null hypothesis is supported or falsified.**

In step 4, you stated a null hypothesis, and predicted what would happen if there is no treatment effect. While the null hypothesis may seem negative and uninteresting, it is important because most statistical techniques can only test a null hypothesis. Therefore, you will probably end up concluding either:

"the data allowed us to reject the null hypothesis" (meaning that there *was* a treatment effect), or
"the data did not allow us to reject the null hypothesis" (meaning that there was no evidence of a treatment effect).

We *don't* say, "We proved there was an effect," or "We proved there was no effect." Although it is common to talk about "experimental proof," experiments do not prove anything. Experiments can only offer evidence than either supports or fails to support hypotheses.

8. **Start the process again with a more refined question about the system.**

For example, assume that the combination of drugs was more effective than either drug taken individually. Next, you might ask if a combination of drugs X, Y, and Z is more effective than a combination of just X and Y. The usual result of an experiment is more questions.

This exercise will allow you to devise and analyze an experiment in which you test yourself and your classmates to determine your reaction times.
Reaction Time Tests Available

One of the basic features of life is that living things react to stimuli from the environment. This may be as obvious as a frog hopping away as you approach, or as subtle as a plant changing its pattern of hormone secretion in response to increasing day length in the spring.

In this experiment, you will use a computer to measure your reaction time. Once you are familiar with the program, you and the other members of your lab team will devise a reaction time hypothesis and an experiment to test it, and collect and analyze your data.

The reaction time program will present you with a stimulus and then ask you to respond by pressing the spacebar. The stimuli can be either visual or auditory, and simple or complex:

- **X at a known location:** An "X" appears in the middle of the screen.
- **Symbol recognition:** You are given a list of from 1-10 letters. Then letters will appear in the middle of the screen, but you press the spacebar only if the letter is on your list. For example, if your list is "A G B," you press the spacebar if you see an A, a G, or a B, but *not* if you see any other letter.
- **Spot the dot:** A period appears somewhere on the screen. The period may either be high-contrast (white on a blue background) or low-contrast (black on a blue background). You may also change the size of the period.
- **Sound:** The computer sounds a tone.
- **Tone recognition:** The computer sounds either a low tone or a high tone. You press the spacebar only if the *high* tone sounds.

These tests will allow a great variety of experiments. For example:

Table 3.1. Possible types of experiments

Type of Experiment	Examples
the same people on *different tests*	reaction time on X at a known location vs. spot the dot, spot the dot with high contrast vs. low contrast
same people, same test, but *before and* experimental treatment	tone recognition *with and without after* some people talking in the background, or reaction to sound *before and after* drinking caffeine
different groups of people on the same test	men vs. women, athletes vs. non-athletes, people who wear glasses vs. people who don't

The list of possible experiments is endless, but we want you to perform a *valid* experiment by following the steps of the scientific method. Two considerations that will help you draw valid conclusions are asking if your experimental design is sound, and determining if your results are statistically significant.

Good Experimental Design

Experimental design includes questions such as how many observations you will use, the order in which the observations will be made, etc. The suggestions below will help you design valid experiments. Remember that a treatment is a test group. It might be different groups of individuals (e.g., men and women), or it might be different tests performed on the same individuals (e.g., spot the dot with high and low contrast), or it might be the same individuals before and after some manipulation (e.g., before and after exercise).

1. **Use adequate replication**. In general, if you want to generalize your results to a larger group, you should have at least 10 different people in each treatment (when testing one group against another), or at least 10 people (when doing a "before and after" test on the same individuals). This will probably require you to go outside your lab group to get enough people. If you *can't* get 10 people, you will have to note that small sample size has weakened the conclusions of your experiment. Perhaps you will only be able to draw conclusions about the individuals tested, not broader groups like all men and all women. Regardless of the number of people, collect at least 10 reaction times per person per treatment. For example, if we were testing men vs. women, we would have each person do 10 reaction times. If we were testing the effects of drinking caffeine, we would have each person do 10 reaction times before caffeine and 10 after it.

2. **You can't make up for a small number of individuals by having each person perform more tests.** All this will yield is ever-more-precise estimates of the reaction times of *these particular* individuals. If Bob is the only male in a male vs. female experiment, and Bob does 100 reaction time tests, this doesn't make Bob any more representative of males as a group.

3. **Avoid bias.** Bias occurs when one treatment has an advantage or disadvantage that has nothing to do with the independent variable. For example, in our male-female experiment, let's say that all the males were athletes, and athletes have faster reaction times. A common kind of bias relates to the time the tests are done. For example, say that in our male-female study that all the females go first and then all the males do the test. If the males watch the females, they may learn tricks that will improve their own performance. One way to combat this problem would be to have males and females alternate as they do the tests.

The final experimental skill you will require will be some knowledge of how scientists use statistics to help them draw conclusions.

The Role of Statistics

Say that you took the same reaction time test ten times, then another ten times, and then another ten times. There is no reason to believe that your reaction time is changing, but it would be very unusual if all three average reaction times came out exactly the same. That is, even the in the absence of any true treatment differences (indeed, even when there *are* no treatments), successive groups of trials will have different means just due to random variation. This random variation will always be present, so how do we distinguish between it and variation caused by

true treatment differences? This is the purpose of experimental statistics.

You may recall that the null hypothesis for any experiment is that there is no difference between the treatments. This is useful because it is easily testable by statistics. In fact, almost all statistics test the null hypothesis by giving us the probability of the observed difference between treatments *if there were really no treatment effects.*   Put another way, statistics gives the probability that the results are due to chance and not some real difference between the treatments. Therefore,

> **if** the computed probability is *low* (usually less than 0.05), we can reject the null hypothesis and accept the alternate hypothesis that there is a treatment effect;

> **if** the computed probability is *high* (usually greater than 0.05), we cannot reject the null hypothesis, and do not have the evidence to declare that there is a treatment effect.

These principles can be seen in the statistical test we will use in this lab, the chi-square (χ^2) median test. This test is easy to understand and can be used in almost any situation in which two (or more) treatments are being compared. Its only drawback is that it can't detect small treatment differences. Statisticians say it is "robust but insensitive."

The Unpaired Chi-Square Median Test

Imagine that two people, A and B, compare their reaction times. They each do 10 tests. Their results (in seconds) are as follows:

Table 3.2. Reaction times of hypothetical individuals A and B.

A: 0.24 0.28 0.32 0.44 0.21 0.19 0.22 0.26 0.20 0.17   average = 0.253

B: 0.23 0.29 0.33 0.20 0.54 0.19 0.23 0.24 0.18 0.25   average = 0.268

Say that we pool these observations in one data set, order them from the fastest time to the slowest time (irrespective of treatment), and then code A observations as an "A" and B observations as a "B." The fastest time was 0.17 (from A), so the first letter on our list should be an A. The next fastest was 0.18 (from B), so the next letter will be a B. After ordering all the observations, we have:

median

fastest A B A B A B A A B B A B B A A B A B A B slowest
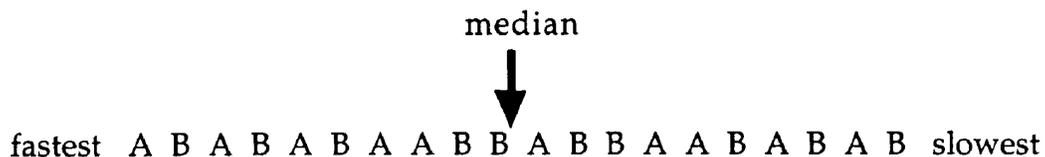
Figure 3.1. The observations in Table 3.2, coded for individuals A and B, and listed from the fastest time to the slowest.

The median divides the data set into two parts. Half the observations will always be below the median and half above it. If we look at the number of *As and Bs* above and below the median, we see:

Table 3.3. The number of A and B reaction times above and below the median.

	A	B
Below Median	5	5
Above Median	5	5

This is exactly the distribution above and below the median we would expect if the two people had the same reaction times.

On the other hand, let's say that now person A compares his reaction time with person C:

Table 3.4. Reaction times of hypothetical individuals A and C.

A:	0.24 0.28 0.32 0.44 0.21 0.19 0.22 0.26 0.20 0.17	average = 0.253
C:	0.15 0.19 0.19 0.16 0.16 0.17 0.15 0.16 0.40 0.25	average = 0.198

This time the coded list is as follows:

median

fastest C C C C C C A C A C A A A A C A A A C A slowest
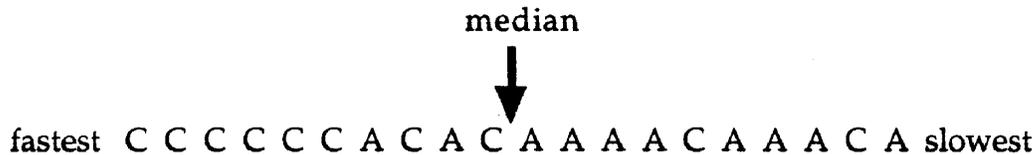
Figure 3.2. The observations in Table 4, coded for individuals A and C, and listed from the fastest time to the slowest.

The table looks like this:

Table 3.5. The number of A and C reaction times above and below the median.

	A	C
Below Median	2	8
Above Median	8	2

If there were really no difference between the treatments, this very uneven distribution, with almost all A observations being slower than the median time and almost all C observations being faster, would be very improbable.

A statistic called chi-square (χ^2) can attach a probability to both Table 3.5 and Table 3.3 above. χ^2 contrasts the counts of observations in classes (for example, above and below the median) with the counts expected if there were really no difference between the treatments. Examples of χ^2 calculations are presented in Appendix A. Statisticians have compiled extensive tables of how often χ^2 values of various sizes occur *in simulated data where there is no difference between the treatments.*

The χ^2 and probability values associated with increasingly uneven distributions of two treatments with ten observations each appear in Table 3.6. The probabilities shown are the probabilities that this distribution above and below the median could have originated just due to chance, not due to any real difference between the treatments. Note how the distributions

become more and more improbable as they become less even.

Table 3.6. χ^2 and probability values associated with increasingly uneven distributions of observations above and below the median in Treatments X and Y.

	X	Y	
Below Median	5	5	$\chi^2 = 0$
Above Median	5	5	Probability > 0.9
Below Median	6	4	$\chi^2 = 0.8$
Above Median	4	6	Probability 0.25-0.50
Below Median	7	3	$\chi^2 = 3.2$
Above Median	3	7	Probability 0.05-0.10
Below Median	8	2	$\chi^2 = 7.2$
Above Median	2	8	Probability 0.005-0.01

In biology, it is traditional to reject the null hypothesis if the observed distribution of data would occur less than 5% of the time if there were no difference between treatments. Therefore, 5% (or 0.05) is called the "critical value." In the A-B experiment, where the probability was greater than 0.90, we would conclude that we can't reject the null hypothesis, and that we have no reason to doubt that A and B have the same reaction time. We would say that the results are *not significant* at the 0.05 level. On the other hand, the null hypothesis is very improbable (probability only 0.005-0.01) in the A-C experiment. We would say that the results were *significant* at the 0.05 level. We *do* have enough evidence to reject the null hypothesis there, and we can accept the alternate hypothesis that there is a difference.

The *unpaired* median test is appropriate where we are testing different individuals or groups of individuals against each other. Where we are testing the same individuals (usually in a before-and-after experiment), a *paired* median test will be more sensitive.

The Paired Median Test

Say that two individuals (Albert and Zelda) in a lab group test their reaction time before and after drinking a soft drink with lots of caffeine. They each do 10 tests before and 10 tests after the soft drink. Their results appear below.

Table 3.7. Reaction times of Albert (slow) and Zelda (fast) before (B) and after (A) drinking a

soft drink with a high caffeine content.

Albert		Zelda	
B	**A**	**B**	**A**
0.43	0.40	0.11	0.09
0.48	0.45	0.13	0.14
0.62	0.50	0.15	0.14
0.51	0.49	0.10	0.11
0.70	0.55	0.14	0.10
0.60	0.64	0.19	0.18
0.55	0.43	0.22	0.25
0.60	0.73	0.24	0.15
0.50	0.45	0.22	0.19
0.43	0.39	0.19	0.11

Did caffeine have an effect? The 20 fastest observations are all Zelda's, and they include an equal mix of befores and afters. Likewise, the 20 slowest observations are all Albert's, and also are an even mix of befores and afters.

Therefore, the final table will be:

Table 3.8. An *unpaired* median test of the reaction times in Table 7.

	Before	**After**	
Above Median	10	10	
Below Median	10	10	$\chi^2 = 0.$

However, note that there is tremendous variation between pitifully slow Albert and lightning fast Zelda, and the test above never took that into account.

Instead of asking if *all* the "befores" are different from *all* the "afters," it would be more appropriate to ask if drinking the soft drink causes a change in reaction time *within the same individual*. This way, variation between individuals won't matter. Let's re-code the data above as follows: if the "after" time is slower than the corresponding "before" time, we'll put down a +; if the "after" time is faster, we'll put down a -. The table then looks like this:

Table 3.9. The change in Table 7's reaction times from "before" to "after" *within the same individual*.

Albert	-	-	-	-	-	+	-	+	-	-
Zelda	-	+	-	+	-	-	+	-	-	-

If caffeine made no difference, we would expect that there would be 10 + and 10 - signs in the table above because there would be no tendency for the "afters" to be higher or lower than the "befores." We would expect that the "befores" would be higher 10 times and the "afters" would be higher 10 times. Instead, we see 5 + and 15 -. When we evaluate observed vs. expected with χ^2, we find that $\chi^2 = 5$, and the probability of a χ^2 this high would be less than 0.025 if there were no difference between the treatments. Using a paired test when appropriate

can make a big difference in your conclusions.

To summarize, an *unpaired* test compares all observations in treatment 1 with all observations in treatment 2. An unpaired test is appropriate when there is no reason to link particular observations in treatment 1 with particular observations in treatment 2. An example would be a male vs. female experiment. On the other hand, a *paired* test compares the first observation in treatment 1 with the first observation in treatment 2, the second observation in treatment 1 with the second observation in treatment 2, etc. It will greatly increase the power of the test in "before and after" experiments where the effect of treatments on *the same individuals* are being examined.

A rule that will work in most cases is:
treatments use *different individuals*—unpaired test;
treatments use *the same individuals*—paired test.

Nonsignificant Results

One final thing: Students are usually disappointed if their experiment does not show a significant treatment effect, but this concern is unnecessary. If a well-designed experiment does not disprove the null hypothesis, we have still found out reliable information about nature.
Non-significance does not imply *insignificance*.

Procedure

1. Turn on your computer. If you're using a Windows machine, your machine should be Windows, not DOS.

2. If you are using a Macintosh, open the "BioBytes" folder, double-click on the "BioBytes" icon, and choose the *Reaction Time* program from the menu. If you are using a Windows machine, select Start/Programs/BioBytes. Then select the *Reaction Time* program from the BioBytes menu.

3. Choose "Collect some reaction time data" from the main menu.

4. Indicate for now that you want to use just the reaction time program.

5. Now look at each of the possible experiments offered by following the directions on the screen. Be thinking about what experiment you and your group want to perform. There is no need to do lengthy experiments here—3 or 4 reaction times per type of test will be fine.

6. When you have looked at all the tests, decide on a test (or tests) you will use, and design your experiment using the seven steps of the scientific method and the principles of experimental design cited at the beginning of this exercise. Consult your instructor if necessary. Fill in the following blanks:
    Your research hypothesis will usually be of the form, "X affects reaction time."
    Research Hypothesis: _____
    _____
    Prediction of Outcome Using the Research Hypothesis: _____
    _____

_____

Your null hypothesis will usually be of the form, "X does not affect reaction time."

Null Hypothesis: _____

_____

Prediction of Outcome Using the Null Hypothesis: _____

_____

_____

Independent Variable: _____

Dependent Variable: _____

Treatment 1 _____

Treatment 2 _____

Treatment 3 _____

Treatment 4 _____

Treatment 5 _____

       (You only need 2 treatments)

Paired or Unpaired Test? _____

(If unsure, consult with your instructor.)

7.   Return to the reaction time menu screen that asked you if you wanted to use just the reaction time program. Indicate that you want to use the reaction time program and then perform an immediate statistical analysis.

8.   Type in how many treatments you will be using (usually 2, but certainly more than 1) and press Enter. If you plan a paired test, you can only use two treatments.

9.   Indicate whether you're going to use a paired or an unpaired test. You should only use a paired test if you have two treatments and will be using the same individuals in the same order, with the same number of observations in each treatment.

10.   As you set up each test, you will be asked to what treatments this group of observations will belong. Be careful to enter this information correctly. There is no need to do treatment 1 first and then treatment 2. If you want to do treatment 2 first, type in that the treatment will be 2 as you set up your first experiment.

11.   As you finish each set of tests, you will be asked if you want to add the observations to its treatment. If the data were valid and do belong in that treatment, indicate yes. If you say no, the most recent group of data will be discarded and will have to be done over. Then the next group of tests will begin. There is no need to write down the results of each test.

12.   **Don't** indicate that this is the end of the experiment unless everyone has completed all their tests and you want the final statistical analysis.

13.   When all the data have been added, the program will send them for statistical analysis. Fill out the Table 3.10 and then *either* Table 3.11 or 3.12 with the results:

Table 3.10. Treatment averages.

Treatment	Average
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____

(You only need two treatments)

Table 3.11. Statistical results for an *unpaired* median test.

Treatment	Below Median	Above Median
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
Chi-square = _____		**Probability** = _____

Accept or reject the null hypothesis? _____

Interpretation _____

_____

Table 3.12. Statistical results for a *paired* median test.

Treatment 1	Treatment 2
Higher	**Higher**
_____	_____
Chi-square = _____	Probability = _____

Accept or reject the null hypothesis? _____

Interpretation _____

_____

Report

  Your instructor may require you to write a report on your experiment; follow the directions you are given.

<div align="center">

Acknowledgements

</div>

Jean Dickey, and Edward Pivorun) under which the Reaction Time software was completed. We would also like the acknowledge the help of Dr. David Cain, who assisted with our presentation at the 1998 ABLE meeting.

Literature Cited

Helms, D. R., C. W. Helms, R. J. Kosinski, and J. R. Cummings. 1998. Biology in the laboratory. Third edition. W. H. Freeman and Company, New York. ISBN 07167-3146-0.

Kosinski, R. J. 1998. Student workbook for BioBytes 3.1: Simulations for the biology laboratory on CD-ROM. W. H. Freeman and Company, New York, 96 pages. ISBN 0-7167-3339-0.

Appendix A: Chi-Square Calculations

χ^2 compares observed versus expected counts, and uses the formula

$$\chi^2 = \Sigma[(O - E)^2/E]$$

where O is the observed count, E is the expected count, and the Σ means summation over all classes. For example, in the A-B experiment and the A-C experiment, we have 10 observations per treatment and therefore we expect 5 observations in each treatment to be above the median and 5 to be below it. There are four classes (treatment A above and below the median, and the other treatment above and below the median). Therefore, for the A vs. B experiment, χ^2 would be

$$\chi^2 = (5 - 5)^2/5 + (5 - 5)^2/5 + (5 - 5)^2/5 + (5 - 5)^2/5 = 0$$

For the A vs. C experiment, χ^2 would be

$$\chi^2 = (2 - 5)^2/5 + (8 - 5)^2/5 + (2 - 5)^2/5 + (8 - 5)^2/5 = 7.2$$

By the way, students will be relieved to learn that they will not have to do any χ^2 calculations. They will automatically be done for them when they log their reaction times.

Appendix B: Sample Student Investigations

The list below shows a sample of the range of student reaction time investigations that have been proposed and used at Clemson:

> men vs. women
> preferred vs. nonpreferred hand
> before and after drinking caffeine
> with and without distraction from music or people talking
> with either rock or classical music played through earphones
> high vs. low contrast in spot-the-dot
> large vs. small symbol recognition list
> visual vs. auditory stimulus
> simple tone stimulus vs. tone discrimination stimulus

Some of these (men vs. women and caffeine) are tried again and again, but usually do not show a significant difference. Others (high vs. low contrast in spot-the-dot, simple tone vs. tone discrimination, auditory vs. visual stimulus, with and without distraction) reliably produce significant differences.

If students do not achieve a significant difference, they commonly think the experiment has been

a failure and must be done over. We have to convince them that a valid finding of no difference between treatments is just as worthwhile as a valid finding of a treatment difference.

Appendix C: A Literature Review on Reaction Time

To help students write their reports, we have posted the information below on a website: http://biology.clemson.edu/bpc/bp/Lab/110/reaction.htm.

Reaction time has been a favorite subject of experimental psychologists since the middle of the nineteenth century. However, most studies ask questions about the organization of the brain, so the authors spend a lot of time trying to determine if the results conform to some mathematical model of brain activity. This makes these papers hard to understand for the beginning student. In this review, I have ignored these brain organization questions and summarized the major literature conclusions that are applicable to undergraduate laboratories using my *Reaction Time* software.

I hope this review helps you write a good report on your reaction time experiment. I also apologize to reaction time researchers for omissions and oversimplifications.

Some Initial Terminology

Psychologists have named three basic kinds of reaction time experiments (Luce, 1986; Welford, 1980):

In *simple* reaction time experiments, there is only one stimulus and one response. "X at a known location," "spot the dot," and "reaction to sound" all measure simple reaction time.

In *recognition* reaction time experiments, there are some stimuli that should be responded to (the "memory set"), and others that should get no response (the "distracter set"). There is still only one correct response. "Symbol recognition" and "tone recognition" are both recognition experiments.

In *choice* reaction time experiments, the user must give a response that corresponds to the stimulus, such as pressing a key corresponding to a letter if the letter appears on the screen. The *Reaction Time* program does not use this type of experiment because the response is always pressing the spacebar.

By the way, professional psychologists doing these experiments typically employ about 20 people doing 100-200 reaction times each...per treatment (Luce, 1986, Ch. 6)! Our experiments of 3 or 4 people doing 10 reaction times each are very small.

Mean Reaction Times

For about 100 years, the accepted figures for mean simple reaction times for college-age individuals have been about 190 ms (0.19 sec) for light stimuli and about 160 ms for sound stimuli (Brebner and Welford, 1980; Fieandt *et al.*, 1956; Galton, 1899; Welford, 1980).

Simple vs. Recognition vs. Choice Reaction Times

The pioneer reaction time study was that of Donders (1868). He showed that a simple reaction time is shorter than a choice reaction time, and that the recognition reaction time is longest of all. Laming (1968) concluded that simple reaction times averaged 220 msec but recognition reaction times averaged 384 msec. This is in line with many studies concluding that a complex stimulus (e.g., several letters in symbol recognition vs. one letter) elicits a slower reaction time (Brebner and Welford, 1980; Luce, 1986; Teichner and Krebs, 1974). An example very much like our experiment was reported by Surwillo (1973), in which reaction was faster when a single tone sounded than when either a high or a low tone sounded and the subject was supposed to react only when the high tone sounded.

Number of Possible Valid Stimuli

Several investigators have looked at the effect of increasing the number of possible stimuli in

recognition and choice experiments. Hick (1952) found that in choice reaction time experiments, response was proportional to log(N), where N is the number of different possible stimuli. In other words, reaction time rises with N, but once N gets large, reaction time no longer increases so much as when N was small. Sternberg (1969) said that in recognition experiments, as the number of items in the memory set increases, the reaction time rises proportionately (that is, proportional to N, not to log N). Reaction times ranged from 420 msec for 1 valid stimulus (such as one letter in symbol recognition) to 630 msec for 6 valid stimuli, increasing by about 40 msec every time another item was added to the memory set. Nickerson (1972) reviewed several recognition studies and agreed with these results.

Light vs. Sound vs. Touch Stimuli

Many researchers have confirmed that reaction to sound is faster than reaction to light, with mean auditory reaction times being 140-160 msec and visual reaction times being 180-200 msec (Brebner and Welford, 1980; Fieandt *et al.*, 1956; Galton, 1899; Welford, 1980; Woodworth and Schlosberg, 1954). Perhaps this is because an auditory stimulus only takes 8-10 msec to reach the brain (Kemp *et al.*, 1973), but a visual stimulus takes 20-40 msec (Marshall *et al.*, 1943). Reaction time to touch is intermediate, at 155 msec (Robinson, 1934).

Brebner and Welford (1980) also cite literature that shows that visual stimuli perceived by different portions of the eye produce different reaction times. The fastest reaction time comes when a stimulus is seen by the cones (when the person is looking right at the stimulus). If the stimulus is picked up by rods (around the edge of the eye), the reaction is slower.

Stimulus Intensity

Froeberg (1907) found that visual stimuli that are longer in duration elicit faster reaction times, and Wells (1913) got the same result for auditory stimuli.

Piéron (1920) and Luce (1986) reported that the weaker the stimulus (such as a very faint light) is, the longer the reaction time is. However, after the stimulus gets to a certain strength, reaction time becomes constant. In other words, the relationship is:
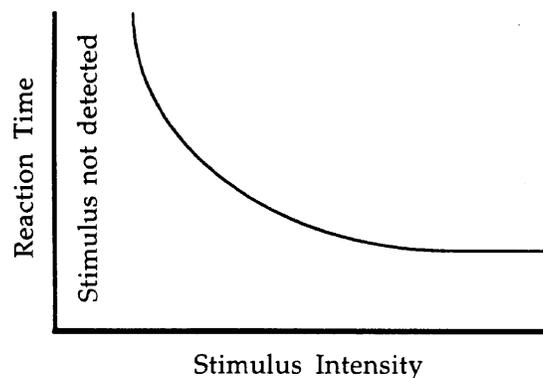
Figure 3.3.
The proposed relation between stimulus intensity and reaction time.

Kohfeld (1971) found that the difference between reaction time to light and sound could be eliminated if a sufficiently high stimulus intensity was used.

Other Factors Influencing Reaction Time

If variation caused by the type of reaction time experiment, type of stimulus, and stimulus intensity are ignored, there are still many factors affecting reaction time.

Arousal. One of the most investigated factors affecting reaction time is "arousal" or state of attention, including muscular tension. Reaction time is fastest with an intermediate level of arousal, and deteriorates when the subject is either too relaxed or too tense (Broadbent, 1971; Freeman, 1933; Welford, 1980). That is, reaction time responds to arousal as follows:
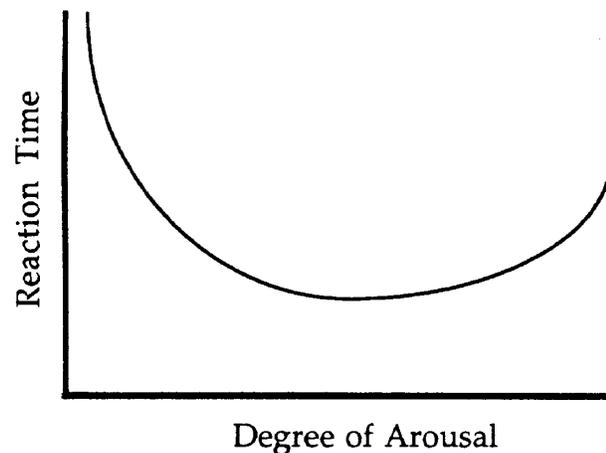
Figure 3.4. The proposed relation between stimulus intensity and reaction time.

Age. Reaction time shortens from childhood into the late 20s, then increases slowly until the 50s and 60s, and then lengthens faster as the person gets into his 70s and beyond (Welford, 1977). An early study (Galton, 1899) reported that for teenagers (15-19) mean reaction times were 187 msec for light stimuli and 158 ms for sound stimuli. Welford (1980) speculates on the reason for slowing reaction time with age. It's not simple mechanical factors like the speed of nervous conduction. It may be the tendency of older people to be more careful and monitor their responses more thoroughly (Botwinick, 1966).

Gender.   At the risk of being politically incorrect, in almost every age group, males have faster reaction times than females, and female disadvantage is not reduced by practice (Noble *et al.*, 1964; Welford, 1980). Bellis (1933) reported that mean time to press a key in response to a light was 220 msec for males and 260 msec for females; for sound the difference was 190 msec (males) to 200 msec (females). In comparison, Engel (1972) reported a reaction time to sound of 227 msec (male) to 242 msec (female). Botwinick and Thompson (1966) found that almost all of the male-female difference was accounted for by the lag between the presentation of the stimulus and the beginning of muscle contraction. Muscle contraction times were the same for males and females.

Fatigue. Welford (1968, 1980) found that reaction time gets slower when the subject is fatigued. Singleton (1953) observed that this deterioration due to fatigue is more marked when the reaction time task is complicated than when it is simple. Mental fatigue, especially sleepiness, has the greatest effect. Kroll (1973) found no effect of purely muscular fatigue on reaction time.

Distraction.   Welford (1980) and Broadbent (1971) reviewed studies showing that

distractions increase reaction time.

Warnings. Brebner and Welford (1980) report that reaction times are faster when the subject has been warned that a stimulus will arrive soon. In the *Reaction Time* program, the delay is never more than about 3 sec, but these authors report that even giving 5 minutes of warning helps.

Order of Presentation. Welford (1980) and Laming (1968) observed that when there are several types of stimuli, reaction time will be faster where there is a "run" of several identical stimuli than when the different types of stimuli appear in mixed order.

Breathing Cycle. Buchsbaum and Calloway (1965) found that reaction time was faster when the stimulus occurred during expiration than during inspiration.

Finger Tremors. Brebner and Welford (1980) report that fingers tremble up and down at the rate of 8-10 cycles/sec, and reaction times are faster if the reaction occurs when the finger is already on the "downswing" part of the tremor.

Personality. Brebner (1980) found that extroverted personality types had faster reaction times, and Welford (1980) and Nettelbeck (1973) said that anxious personality types had faster reaction times.

Exercise. Exercise can affect reaction time. Welford (1980) found that physically fit subjects had faster reaction times, and both Levitt and Gutin (1971) and Sjoberg (1975) showed that subjects had the fastest reaction times when they were exercising sufficiently to produce a heart rate of 115 beats per minute.

Punishment. Shocking a subject when he reacts slowly does shorten reaction time (Johanson, 1922; Weiss, 1965).

Stimulant Drugs. Kleemeier *et al.* (1956) found that administering an amphetamine-like drug to a group of elderly men did not make their reaction times faster, although it did make their physical responses more vigorous.

Intelligence. Serious mental retardation produces slower and more variable reaction times. Among people of normal intelligence, there is a slight tendency for more intelligent people to have faster reaction times, but there is much variation between people of similar intelligence (Nettelbeck, 1980).

Literature Cited

Bellis, C. J. 1933. Reaction time and chronological age. Proceedings of the Society for Experimental Biology and Medicine, 30:801.

Botwinick, J. 1966. Cautiousness in advanced age. Journal of Gerontology, 21:347-353.

Botwinick, J. and L. W. Thompson. 1966. Components of reaction time in relation to age and sex. Journal of Genetic Psychology, 108:175-183.

Brebner, J. T. 1980. Reaction time in personality theory. Pages 309-320, *in* Reaction times (A. T. Welford, Editor). Academic Press, New York, 418 pages. ISBN 0127428801.

Brebner, J. T. and A. T. Welford. 1980. Introduction: an historical background sketch. Pages 1-23, *in* Reaction times (A. T. Welford, Editor). Academic Press, New York, 418 pages. ISBN 0127428801.

Broadbent, D. E. 1971. *Decision and Stress.* Academic Press, London, 522 pages. ISBN 0121355500.

Buchsbaum, M. and E. Callaway. 1965. Influence of respiratory cycle on simple RT. Perceptual and Motor Skills, 20:961-966.

Donders, F. C. 1868. On the speed of mental processes. Translated by W. G. Koster, 1969.

Acta Psychologica, 30:412-431.

Engel, B. T., P. R. Thorne, and R. E. Quilter. 1972. On the relationship among sex, age, response mode, cardiac cycle phase, breathing cycle phase, and simple reaction time. Journal of Gerontology, 27:456-460.

Fieandt, K. von, A. Huhtala, P. Kullberg, and K. Saarl. 1956. Personal tempo and phenomenal time at different age levels. Reports from the Psychological Institute, No. 2, University of Helsinki.

Freeman, G. L. 1933. The facilitative and inhibitory effects of muscular tension upon performance. American Journal of Psychology, 26:602-608.

Froeberg, S. 1907. The relation between the magnitude of stimulus and the time of reaction. Archives of Psychology, No. 8.

Galton, F. 1899. On instruments for (1) testing perception of differences of tint and for (2) determining reaction time. Journal of the Anthropological Institute, 19:27-29.

Hick, W. E. 1952. On the rate of gain of information. Quarterly Journal of Experimental Psychology, 4:11-26.

Johanson, A. M. 1922. The influence of incentive and punishment upon reaction-time. Archives of Psychology, No. 54.

Kohfeld, D. L. 1971. Simple reaction time as a function of stimulus intensity in decibels of light and sound. Journal of Experimental Psychology, 88:251-257.

Kemp, B. J. 1973. Reaction time of young and elderly subjects in relation to perceptual deprivation and signal-on versus signal-off condition. Developmental Psychology, 8:268-272.

Kleemeier, R. W., T. A. Rich, and W. A. Justiss. 1956. The effects of alpha-(2-piperidyl) benzhydrol hydrochloride (Meratran) on psychomotor performance in a group of aged males. Journal of Gerontology, 11:165-170.

Kroll, W. 1973. Effects of local muscular fatigue due to isotonic and isometric exercise upon fractionated reaction time components. Journal of Motor Behavior, 5:81-93.

Laming, D. R. J. 1968. Information theory of choice-reaction times. Academic Press, London, 172 pages.

Levitt, S. and B. Gutin. 1971. Multiple choice reaction time and movement time during physical exertion. Research Quarterly, 42:405-410.

Luce, R. D. 1986. Response times: their role in inferring elementary mental organization. Oxford University Press, New York, 562 pages. ISBN 0195036425.

Marshall, W. H., S. A. Talbot, and H. W. Ades. 1943. Cortical response of the anaesthetized cat to gross photic and electrical afferent stimulation. Journal of Nerophysiology, 6:1-15.

Nettelbeck, T. 1973. Individual differences in noise and associated perceptual indices of performance. Perception, 2:11-21.

Nettelbeck, T. 1980. Factors affecting reaction time: Mental retardation, brain damage, and other psychopathologies. Pages 355-401, *in* Reaction times (A. T. Welford, Editor). Academic Press, New York, 418 pages. ISBN 0127428801.

Nickerson, R. S. 1972. Binary-classification reaction times: A review of some studies of human information-processing capabilities. Psychonomic Monograph Supplements, 4:275-318.

Noble, C. E., B. L. Baker, and T. A. Jones. 1964. Age and sex parameters in psychomotor learning. Perceptual and Motor Skills, 19:935-945.

Piéron, H. 1920. Nouvelles recherches sur l'analyse du temps de latence sensorielle et sur la loi qui relie ce temps a l'intensité de l'excitation. Année Psychologique, 22:58-142.

Robinson, E. S. 1934. Work of the integrated organism. *In* Handbook of general experimental psychology (C. Murchison, Editor). Clark University Press, Worcester, MA.

Singleton, W. T. 1953. Deterioration of performance on a short-term perceptual-motor task. Pages 163-172 *in* Symposium on fatigue (W. F. Floyd and A. T. Welford, Editors). H. K. Lewis and Co., London.

Sjoberg, H. 1975. Relations between heart rate, reaction speed, and subjective effort at different work loads on a bicycle ergometer. Journal of Human Stress, 1: 21-27.

Sternberg, S. 1969. Memory scanning: Mental processes revealed by reaction time experiments. American Scientist, 57:421-457.

Surwillo, W. W. 1973. Choice reaction time and speed of information processing in old age. Perceptual and Motor Skills, 36:321-322.

Teichner, W. H. and M. J. Krebs. 1974. Laws of visual choice reaction time. Psychological Review, 81:75-98.

Weiss, A. D. 1965. The locus of reaction time change with set, motivation, and age. Journal of Gerontology, 20:60-64.

Welford, A. T. 1968. Fundamentals of skill. Methuen, London, 426 pages. ISBN 0416030009.

Welford, A. T. 1977. Motor performance. Pages 450-496 *in* Handbook of the psychology of aging (J. E. Birren and K. W. Schaie, Editors). Van Nostrand Reinhold, New York.

Welford, A. T. 1980. Choice reaction times: Basic concepts. Pages 73-128 *in* Reaction times (A. T. Welford, Editor). Academic Press, New York, 418 pages. ISBN 0127428801.

Wells, G. R. 1913. The influence of stimulus duration on RT. Psychological Monographs, 15:1066.

Woodworth, R. S. and H. Schlosberg. 1954. Experimental psychology. Henry Holt, New York, 948 pages.