

Chapter 11

The Scientific Method: An Introduction Using Reaction Time

Bob Kosinski and John Cummings

Division of Life Science Studies
Genetics, Biochemistry, and Life Science Studies
Clemson University
Clemson, SC 29634-0318
Voice: (864) 656-3830; Fax: (864) 656-3839
rjksn@clemson.edu
cumminj@clemson.edu

Bob Kosinski received his B.S. 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 the role of investigative laboratories, computer simulations, and physiological interfacing in introductory biology, plus collaboration between science education and agricultural education.

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.

Reprinted From: Kosinski, B. and J. Cummings. 2004. The scientific method: An introduction using reaction time. Pages 219-234, *in* Tested studies for laboratory teaching, Volume 25 (M. A. O'Donnell, Editor). Proceedings of the 25th Workshop/Conference of the Association for Biology Laboratory Education (ABLE), 414 pages.

- Copyright policy: <http://www.zoo.utoronto.ca/able/volumes/copyright.htm>

Although the laboratory exercises in ABLE proceedings volumes have been tested and due consideration has been given to safety, individuals performing these exercises must assume all responsibility for risk. The Association for Biology Laboratory Education (ABLE) disclaims any liability with regards to safety in connection with the use of the exercises in its proceedings volumes.

© 2004 W. H. Freeman and Company

Contents

Notes for Instructor	
The <i>Reaction Time</i> Software	220
Using the Exercise	221
Difficulties Experienced	222
Materials	222
Acknowledgements.....	223
Literature Cited.....	223
Student Outline	
Steps of the Scientific Method.....	224
The Reaction Time Software	225
Good Experimental Design.....	226
The Role of Statistics	226
Report.....	230
Procedure	231
Appendix A: χ Calculations.....	234

Notes for Instructor

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 exercise plus its introduction beforehand takes about two hours to complete.

The *Reaction Time* Software.

The only equipment the exercise requires is a Windows-compatible or Macintosh computer and Kosinski's *Reaction Time* program. Windows and Macintosh versions of this program have been published by W. H. Freeman and Company (Kosinski, 1999) and at the time of this writing were available for download at <<http://www.whfreeman.com/biolab/helms/biobytes.html>>. Copyright to the program, and the student exercise that follows, is owned by W. H. Freeman. The student exercise is used in this compilation with the permission of W. H. Freeman.

The *Reaction Time* 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 intentionally elementary, graduate student

teaching assistants have told us that it helped them understand their high-level statistics courses for the first time.

Using the Exercise

One computer per four to six students works well. Because all the data used in one experiment must be put entered into one computer, the group at one computer cannot be too small or it will not be able to produce enough observations. Also, a larger group leads to a better discussion. The exercise takes about two hours; the most time-consuming part is the explanation of the scientific method and statistics.

The experiments the students could propose are infinitely varied, but can be categorized into several types:

1. testing one group of people against another (e.g., men vs. women, tall people and small people) on the same reaction time test;
2. comparing the reaction times of the same people on two different reaction time tests (such “spot-the-dot” with high- vs. low-contrast backgrounds);
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 on the World Wide Web at <http://biology.clemson.edu/bpc/bp/Lab/110/reaction.htm>. As of 2003, Kosinski was updating this literature review annually.

We have used this exercise for several years in both our majors and our nonmajors freshman general biology courses at Clemson University. Student reaction 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 from the start. Students have no difficulty using the software or devising investigations, although at times they want us to tell them the “right” investigation to do.

Common Student Investigations

Students tend to select the same types of experiments again and again. Some common investigations that usually don't produce significant results are: males vs. females, blonds (thought to be slower) vs. brunettes, pressing the spacebar with the dominant vs. non-dominant hand, with and without distraction, athletes vs. non-athletes, and southerners (thought to be slower) vs. northerners. Traditional investigations that *do* produce significant results are “X at a Known Location” vs. symbol recognition, high-contrast vs. low-contrast in “Spot-the-Dot,” short vs. long lists of letters in Symbol Recognition, “Reaction to Sound” vs. “Tone Recognition,” and auditory stimuli vs. visual stimuli. In general, experiments comparing different groups of people tend not to produce significantly different results, but different reaction time tests usually do produce significantly different reaction times.

Difficulties Experienced

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.

Students sometimes want to test humorous hypotheses (like the old favorite, blonds vs. brunets), but we must remind them that they must produce a biological rationale for their hypothesis.

An aspect of the lab that may be difficult for *instructors* is the distinction made between a scientific (or explanatory) hypothesis and a statistical hypothesis. Scientists are familiar with the idea of a hypothesis as a provisional explanation of the way a system works, and with the use of hypotheses to make predictions that are tested by experiment. We could also regard an explanatory hypothesis as the biological rationale for an experiment.

However, in statistics, hypotheses are not explanations, but rather statements that there either is or is not a difference between two or more variables. For example, we test the hypothesis that the mean of treatment A and the mean of treatment B are the same (really, that the difference between them is equal to zero). Because this difference must either be equal to zero or *not* equal to zero, the null and research hypotheses are mutually exclusive, and to reject one is to fail to reject the other.

Therefore, we have students write out explanatory hypotheses and make predictions from them, and then we ask them separately about their research and null hypotheses. For example:

Explanatory Hypothesis: Taller people have longer nervous pathways than shorter people. Also, action potentials travel at a finite speed.

Prediction: Tall people will have slower reaction times than short people.

Research Hypothesis: The reaction times of tall people and short people are different.

Null Hypothesis: The reaction times of tall people and short people are the same.

Materials

The only materials required are Windows-compatible or Macintosh computers and the *Reaction Time* software. Windows and Macintosh versions of the Kosinski reaction time program (plus five other Kosinski programs) 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). The "BioBytes" CD (Kosinski, 1999) is available for \$12.95 from W. H. Freeman:

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

Also, the reaction time program alone is available for free download at <http://www.whfreeman.com/biolab/helms/biobytes.html>. Although this Web site refers to Windows 95, the Windows program will work with Windows 95, 98, 2000, and XP. The Macintosh version will work with OS versions ranging at least from 8.1 to 10.2.3.

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). Usually, the room is too noisy to use audio stimuli without earphones.

Acknowledgements

We gratefully acknowledge the support of the Course and Curriculum Section of the National Science Foundation, which provided funding (Project 9156246, to Robert Kosinski, Jean Dickey, and Edward Pivorun) under which the *Reaction Time* software was completed. We also acknowledge W. H. Freeman's permission to use the exercise in this ABLE compilation.

Literature Cited in Instructor's Introduction

- 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. 1999. 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.

Online Literature Review on Reaction Time

<<http://biology.clemson.edu/bpc/bp/Lab/110/reaction.htm>>

Student Outline

The Scientific Method

The scientific approach is a powerful method for understanding the natural world because it is founded on our observations of how the world works. To be successful, the observations must be systematic and objective, and must be analyzed according to certain rules.

This exercise will teach you the scientific method by allowing you and your classmates to devise and analyze an experiment on reaction time. You will use a computer program that will present you with a stimulus and then ask you to respond by pressing the spacebar. There are several types of stimuli you can use, but more on that later. Once you are familiar with the program, you and the other members of your lab team will devise a reaction time experiment and analyze your data.

The Steps of the Scientific Method

1. Use your general and perhaps non-systematic observations to make a generalization.

In talking about reaction time, the males in the group assert that males must have faster reaction times. After all, they play a lot of sports, they like fast driving, so yeah, they've gotta have faster reaction times!

2. Devise one or more explanatory hypotheses to explain the generalization.

An explanatory (or scientific) hypothesis is the reasoning behind the generalization. It is guess about the way the system works that will allow us to make a prediction. In this case: Many typically male activities (such as sports) train males to have faster reaction times. Females engage in few activities that require fast reaction times.

3. Define the experiment's independent, dependent, and standardized variables.

The *independent variable* is the factor that is being tested in the current experiment. In this case, the independent variable is gender because we are testing the idea that males and females have different reaction times.

The *dependent variable* is the aspect of the system that is that is showing response to the independent variable. Here, the dependent variable is reaction time. We think it may be dependent on gender.

The *standardized variables* are all the variables that might influence the dependent variable (reaction time) but are irrelevant to this experiment. The standardized variables are held constant so they will not bias the results. The number of reaction time tests performed, the type of tests, the room conditions, etc., are all standardized variables in this experiment.

4. Use the explanatory hypothesis and the variables to make a falsifiable prediction.

In a fair reaction time comparison, males will have faster reaction times. This is falsifiable because if the females are faster, or the two genders have the same reaction time, we will see that outcome.

5. Use the prediction to devise a null hypothesis and a research hypothesis.

“Hypothesis” is often used in science to mean an educated guess about the way a system works. That's what our explanatory (or scientific) hypothesis was. However, in the statistics used in our experiments, “hypothesis” has a very restricted meaning. The “null” hypothesis is a statement of no difference: “Males and females have the same reaction times.” The “research” hypothesis is the opposite statement: “Males and females have different reaction times.” Note that the

research hypothesis does not say males will be faster or slower. It merely says the reaction times of the two genders will be different. Statistics usually test the null hypothesis. The only way the research hypothesis can be supported is if the null hypothesis is rejected. Also, it is an old scientific custom to “fail to reject” the null hypothesis *unless we have the evidence to reject it*.

6. Define the experiment’s treatments.

A treatment is a group that experiences the same conditions with respect to the independent variable. In this case, the independent variable is gender and the two treatments are males and females. If the experiment were about the effect of exercise on reaction time, the two treatments would be before exercise and after exercise. If the experiment were about the effect of music on reaction time, the two treatments would be with and without music. There can be more than two treatments in an experiment, but most student reaction time experiments use two treatments.

7. Devise an experiment to test the null hypothesis. Try to keep all standardized variables the same for all treatments and avoid bias.

Take an equal number of males and females and have each of them do 10 repetitions of the same reaction time test under identical conditions. So neither treatment will have an advantage, alternate males and females. Because feelings that “our side is winning” (or losing) might influence the results, don’t let anyone know what the results were until all the tests are over.

8. Perform the experiment and collect the data.

9. Analyze the data and determine whether the null hypothesis is rejected or cannot be rejected.

When you use statistical analysis, you will probably end up concluding either:

“*the data allowed us to reject the null hypothesis,*” meaning that there *was* a difference between genders, or

“*the data did not allow us to reject the null hypothesis,*” meaning that there was no evidence of a gender effect, and so we cannot reject the null hypothesis that men and women have the same reaction times.

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 that either supports or fails to support hypotheses.

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

Perhaps we could not reject the null hypothesis, but we did notice that some males who were varsity athletes outperformed the other males and the females. This suggests questions about athlete males vs. nonathlete males, athlete males vs. athlete females, and so forth. The usual outcome of an experiment is more questions.

The Reaction Time Software

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 have either large or small dots appear.
- **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, 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

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., exercise or drinking coffee).

1. **Use adequate replication.** Replication is the number of repeat experiments you do. There are complicated principles governing the number of replicates to use, but for the purposes of this exercise, remember the number 10. 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 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.** If Bob is the only male in a male vs. female experiment, and Bob does 1000 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, 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 Role of Statistics

Say that you measured the reaction times of two groups of people. One group seemed faster than the other, but not by much. So what do you conclude? How big a difference between groups do we need before we reject the null hypothesis? Giving assistance with this problem is the purpose of experimental statistics.

Statistics give us the probability of the observed difference between treatments *if there were really no treatment effects*. Put another way, statistics gives us the probability that the results are due to chance and not some real difference between the treatments. In biology, it is traditional to reject the null hypothesis if this probability is 5% (0.05) or less. We say that 0.05 is our significance

Again, in biology, the traditional significance level is 5% (0.05). We would say the last case above shows a significant difference between X and Y at the 0.05 level. The case above it does not because “0.05-0.10” is not equal to or less than 0.05.

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 (in seconds) appear in Table 4 below.

Table 4. The reaction times of two students before and after drinking a soft drink.

Albert		Zelda	
Before	After	Before	After
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 for an unpaired median test will be:

Table 5. The results of an unpaired median test on “Albert” and “Zelda” above.

	Before	After
Below Median	10	10
Above Median	10	10

$$\chi^2 = 0$$

Probability = 0.75-0.90

This implies that caffeine made no difference. 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” (as in an unpaired median test), 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. This is called a *paired test*. 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 -. Each before-after pair will generate either a + or a -. The table then looks like this:

Table 6. The numbers of before-to-after changes that were decreases (-) or increases (+).

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

If caffeine made no difference, we would expect that there would be 10 + and 10 - signs in Table 6. In other words, the reaction times before caffeine would tend to be faster than the after times half the time, and slower the other half of the time. Instead, we see 5 + and 15 -. There is a strong tendency for the “after” times to be faster. In other words:

Table 7. Expected numbers of + and - changes if before and after times were the same.

	+	-
Pairs	10	10

Table 8. Observed numbers of + and - changes.

	+	-
Pairs	5	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 0.025-0.05 if there were no difference between the treatments. Using a paired test when appropriate can make a big difference in your conclusions.

An *unpaired* test compares all observations in treatment 1 (in no particular order) with all 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.

Statistically non-significant results

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

Non-significance does not imply *insignificance*.

Report

Your instructor will give you directions on how to write a report. You may be helped by a review of the reaction time literature that appears on the World Wide Web at <http://biology.clemson.edu/bpc/bp/Lab/110/reaction.htm>.

Procedure

1. Turn on your computer if it's not on already.
2. If you're using a Windows machine, your machine should be Windows, not DOS. Select Start/Programs/BioBytes. Then select the *Reaction Time* program from the BioBytes menu. 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.
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. Think 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 10 steps of the scientific method and the principles of experimental design cited at the beginning of this exercise. Consult your instructor if necessary. One warning--while you can test humorous hypotheses like, "Blonds are slower than brunets," you will have a terrible time writing your paper because you have to come up with a biological justification for your hypothesis.

Fill in the following blanks:

Explanatory hypothesis _____

Prediction from explanatory hypothesis _____

Now state your statistical hypotheses:

Research Hypothesis: _____

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. From the menu of possible tests, press “Cancel,” indicate you want to collect reaction time data, and indicate that you want to use the reaction time program and then perform an immediate statistical analysis. Press “OK.” **Not doing this is the most common student error with the *Reaction Time* program. You must go back to the first screen and indicate that you want to use the reaction time program and then do 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 treatment 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. There is no need to write down the individual reaction times. You will be given a summary with everything you need at the end.
12. 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.
13. **Don’t** indicate that this is the end of the experiment unless every student has completed all tests and you want the final statistical analysis.
14. When all the data have been added, the program will perform statistical analysis. Fill out the Table 9 and then *either* Table 10 or 11 with the results:

Table 9. Treatment averages.

Treatment	Average

(You only need two treatments.)

Table 10. Statistical results for an *unpaired* median test.

Treatment	Above Median	Below Median

(You only need two treatments.)

Chi-square = _____ Probability = _____

Reject or fail to reject the null hypothesis? _____

Interpretation _____

Table 11. Statistical results for a *paired* median test.

Trt 1 Higher	Trt 2 Higher

Chi-square = _____ Probability = _____

Reject or fail to reject the null hypothesis? _____

Interpretation _____

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 treatment B 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

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

You will be relieved to learn that you will not have to do any χ^2 calculations. They will automatically be done for you when you log your reaction times.