When 95% Accurate Isn't

Todd CadwalladerOlsker
California State University, Fullerton
tcadwall@fullerton.edu

Published: November 2012

Overview of Lesson
In this activity, students will investigate Bayes' theorem using simulated data (generated by a calculator), expected frequencies, and probabilities. By progressing through these perspectives in a deliberate order (first simulated data, then expected frequencies, ending with probabilities) students will "discover" Bayes' theorem for themselves. Along the way, they will be surprised by the role that false positives play in determining conditional probabilities.

GAISE Components
This investigation somewhat follows the four components of statistical problem solving put forth in the Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report. The four components are: formulate a question, design and implement a plan to collect data, analyze the data by measures and graphs, and interpret the results in the context of the original question. This is a GAISE Level B activity.

Common Core State Standards for Mathematical Practice
1. Make sense of problems and persevere in solving them.
2. Reason abstractly and quantitatively.
3. Construct viable arguments and critique the reasoning of others.
7. Look for and make use of structure.
8. Look for and express regularity in repeated reasoning.

Common Core State Standards Grade Level Content (High School)
S-CP. 5. Recognize and explain the concepts of conditional probability and independence in everyday language and everyday situations.
S-CP. 6. Find the conditional probability of A given B as the fraction of B's outcomes that also belong to A, and interpret the answer in terms of the model.

NCTM Principles and Standards for School Mathematics
Data Analysis and Probability Standards for Grades 9-12
Understand and apply basic concepts of probability:
- understand the concepts of sample space and probability distribution and construct sample spaces and distributions in simple cases;
- use simulations to construct empirical probability distributions;
- compute and interpret the expected value of random variables in simple cases;
- understand the concepts of conditional probability and independent events.

Prerequisites
Students will need sufficient knowledge of programmable calculators to run a program to simulate data. Students may need to be able to transfer programs from one calculator to another; alternatively, the instructor can transfer the program to calculators ahead of time. Students will need a basic understanding of expected values and conditional probabilities. Students should *not* have previously seen Bayes' theorem, as they will "discover" Bayes' theorem through this activity.

Learning Targets
Students will be able to use Bayes' theorem to compute conditional probabilities based on given information. Students will be able to re-create Bayes' theorem using expected frequencies.

Time Required
1 class period, possibly 2 if the instructor and students want to investigate further questions.

Materials Required
A copy of the Activity Sheet (pages 13-16); a programmable calculator (TI-83, TI-84, or similar); the program ACPPROG for TI-84 calculators. ACPPROG is available from the STEW website as a compressed .zip file; instructions for transferring the program to TI-84 calculators are included in the .zip file. The ACPPROG code is included at the end of this lesson plan; instructors can adapt it for use in other calculators or programs, such as Excel.

Instructional Lesson Plan

The GAISE Statistical Problem-Solving Procedure

I. Formulate Questions(s)
Begin the lesson with a class discussion of the question, "What does it mean to say that a test, or an observation, is 95% accurate?" The instructor may wish to use specific contextual questions, such as, "What does it mean to say that a weather forecaster is 95% accurate?" or "What does it mean to say that a spam filter is 95% accurate?" This can (and should) lead to some discussion of the meaning of conditional probability statements: for example, the class could discuss the difference between a spam filter that filters 95% of the actual spam and one for which 95% of the filtered messages are spam, and how those two statements could be represented as conditional probabilities. In this example, if S is the event that an email is spam and F is the event that the email is filtered, a spam filter that filters 95% of the spam is one for which $P(F|S) = .95$. On the other hand, a filter for which 95% of the filtered messages are spam is one for which $P(S|F) = .95$.

Next, the instructor should distribute the activity sheet (pages 13-16) and (if necessary) programmable calculators with the ACPPROG program loaded. The class should discuss the following oil exploration problem given at the beginning of the activity sheet:

When exploring an area for oil, oil company surveyors will look for leads, formations on the Earth's surface that indicate the possibility of oil beneath the surface. However, just because it is possible to find oil beneath these leads does not mean that there is oil. Suppose that only 6% of leads actually have oil present. The oil company's surveyors will run a series of tests to determine the likelihood of the presence of oil. The surveyors' tests are 95% accurate for both the presence and the absence of oil: If oil is present, the surveyors' tests will reveal that oil is present with a probability of 0.95, and if oil is absent, the surveyors' tests will reach a negative conclusion with a probability of 0.95. If a lead does, in fact, test positive, the company will drill an exploratory well to determine if oil is actually present. Such exploratory wells are expensive to drill, so the oil company wants to know: what is the probability that a positive-testing lead actually has oil present?

The problem, as stated, gives a very specific meaning to the idea that the tests for oil are 95% accurate. The instructor might have the students work in small groups for the first three questions, or use a whole class discussion. The first three questions on the activity sheet will help students understand the project, and the discussion of conditional probabilities should continue as these questions are answered by the class.

Explain to students that each lead has two events that will be measured, and each event has two possible outcomes: The event of the presence of oil has possible outcomes of "Oil" or "No Oil"; the event of the surveyors' test has possible outcomes of "Positive" or "Negative". Let Event A represent the event that the lead contains oil, and Event B represent the event that the surveyors' test is positive.

Questions 1, 2, and 3; along with some typical student answers in red text are below:

1. What do the following mean, in English?
$P(A)$ The probability that a lead has oil present.
$P(B|A)$ The probability that a lead tests positive, given that there is oil present.
$P(B^C|A^C)$ The probability that a lead tests negative, given that there is not oil present.

2. What are the values of each of these probabilities, according to the problem?
$P(A) = 0.06$
$P(B|A) = 0.95$
$P(B^C|A^C) = 0.95$

3. If the surveyors' test of a lead is positive, what would you guess to be the probability that the lead actually has oil present? Most students' guesses tend to be between 80% and 95%.

Students should share their guesses for the probability that oil is present in question 3.

II. Design and Implement a Plan to Collect the Data
Question 4 asks students to use the ACPPROG program to generate simulated data:

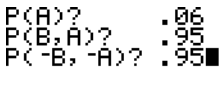
4. We'll estimate the probability of each possible outcome by simulating tests of a large number of leads. Using the program ACPPROG and your values from question 1, generate 1000 data points. ACPPROG uses the –sign to represent the complement, so –A represents A^C, for example. Record the values generated by ACPPROG here:
$(A\&B)$:
$(A\&B^C)$:
$(A^C\&B)$:
$(A^C\&B^C)$:

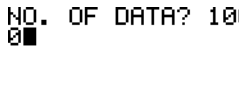
The program will ask for the probability that

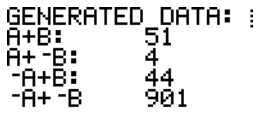
- oil is present in a lead, $P(A)$,
- the probability of a positive test when oil is present, P(B|A), and
- the probability of a negative test when oil is not present, $P(B^C|A^C)$.

Due to limitations of the TI-84 calculator, the symbol | is replaced with a comma, and complements are represented with a – rather than a superscript C. For the oil exploration problem, $P(A) = 0.06$, $P(B|A) = 0.95$, and $P(B^C|A^C) = 0.95$. Students should have found these values in question 2.

The ACPPROG will then ask for the number of data points to generate. Students will ask for 1000 data points. ACPPROG may take some time to run, but should eventually determine a number of data points falling into each of four categories: oil is present and the test is positive $(A\&B)$, oil is present but the test is negative, $(A\&B^C)$, oil is not present but the test is positive, $(A^C\&B)$, and oil is present and the test is negative, $(A^C\&B^C)$. The screenshots below indicate what students should enter, and a typical result from ACPPROG:







These simulated data will be recorded in response to question 4. Each student will have a different set of simulated data, generated using the calculator's random number generating function and the probabilities entered into the calculator. Thus, there will be some variance in students' answers. Student responses based on the particular simulated data in the screenshots above are entered below as typical student answers, in red text:
$(A\&B)$: 51
$(A\&B^C)$: 4
$(A^C\&B)$: 44
$(A^C\&B^C)$: 901

Students will then arrange their results in a table:

	Positive Test	Negative Test
Oil Present	51	4
No Oil Present	44	901

(Note: The sum of all four boxes should be equal to 1000 total leads.)

III. Analyze the Data

Questions 5 through 7 ask students to use their generated data to answer the question asked in question 3: if the test is positive, what is the probability that oil is present? Student responses based on the generated data above are in red:

5. How many leads tested positive for oil? $51 + 44 = 95$

6. Of those leads that tested positive for oil, how many actually have oil present? 51

7. Based on your answers to 5 and 6, what is the probability that a lead has oil present given that it tests positive? In other words, what is $P(A|B)$ based on this simulated data? $\frac{51}{95} = 0.537$

Each student will get slightly different numbers, due to the variance involved in simulated data. Students should compare their results, and the instructor can total the data generated by some or all of the students to get a better estimate of the probability that oil is present. Before continuing, the class should discuss how closely their resulting probability in question 7 matches their guesses in question 3, and whether or not the results are surprising. The instructor may also note that question 7 is answered using $P(A|B) = \frac{n(A\&B)}{n(B)}$, and comment on why this is equivalent to the traditional conditional probability formula, $P(A|B) = \frac{P(A\&B)}{P(B)}$.

The remainder of the activity asks students to use expected values and probabilities instead of simulated data to recalculate the probability that oil is present when the test is positive. The order in which the different methods are used is important: students should first see that the expected frequencies are similar to their simulated data, and should next see that the probabilities are related to their expected frequencies. Here, students should all get the same answers, which are again indicated in red text.

Questions 8 through 14 ask students to use expected frequencies to calculate the probability that oil is present when the test is positive:

8. Suppose the oil company tests 1000 leads. We can expect 6% of those leads to have oil. How many leads should we expect to have oil?
6% of 1000 = 60 leads should have oil.

9. Of those leads that contain oil, how many should we expect to (correctly) test positive for oil, and how many should we expect to (incorrectly) test negative?
95% of 60 = 57 leads will test positive, 5% of 60 = 3 leads will test negative.

10. Of those leads that do not contain oil, how many should we expect to (incorrectly) test positive for oil, and how many should we expect to (correctly) test negative?
5% of 940 = 47 leads will test positive, 95% of 940 = 893 leads will test negative.

11. Summarize your answers to 9 and 10 in the table below:

	Positive Test	Negative Test
Oil Present	57	3
No Oil Present	47	893

(Note: The sum of all four boxes should be equal to 1000 total leads.)

12. How many leads should we expect to test positive for oil? $57 + 47 = 104$

13. Of those leads that we expect to test positive for oil, how many should we expect to actually have oil present? 57

14. Based on your answers to 12 and 13, what is the probability that a lead has oil present given that it tests positive? In other words, what is $P(A|B)$ based on these expected values?
$\frac{57}{104} = 0.548$

Students should see that their expected values in question 11 are similar to their simulated data. If time allows, the instructor can average the students' simulated data to demonstrate that these averages are very close to the expected values. Students will again calculate their answer to question 14 using $P(A|B) = \frac{n(A\&B)}{n(B)}$.

Questions 15 through 21 ask students to calculate the probability that oil is present when the test is positive one last time, using probabilities rather than expected values:

Suppose the oil company tests a large (but unspecified) number of leads.
15. What percentage of those leads to we expect to have oil present? 6%

16. Of those leads that contain oil, we expect 95% to correctly test positive for oil. What percentage of *all* of the leads should we expect to (correctly) test positive, and what percentage of *all* of the leads should we expect to (incorrectly) test negative?
95% of 6%, which gives $0.95 \times 0.06 = 0.057$, or 5.7%. correctly test positive.
5% of 6%, which gives $0.05 \times 0.06 = 0.003$, or 0.3%. incorrectly test negative.

17. Of those leads that do not contain oil, we expect 95% to correctly test negative for oil. What percentage of *all* of the leads should we expect to (incorrectly) test positive, and what percentage of *all* of the leads should we expect to (correctly) test negative?
5% of 94%, which gives $0.05 \times 0.94 = 0.047$, or 4.7%. incorrectly test positive.
95% of 94%, which gives $0.95 \times 0.94 = 0.893$, or 89.3%. correctly test negative.

18. Summarize your answers to 16 and 17 in the table below:

	Positive Test	Negative Test
Oil Present	5.7%	0.3%
No Oil Present	4.7%	89.3%

(Note: The sum of all four boxes should be equal to 100% of the total leads.)

19. What percentage of leads should we expect to test positive for oil?
$5.7\% + 4.7\% = 10.4\%$

20. Of those leads that we expect to test positive for oil, what percentage should we expect to actually have oil present? 5.7%

21. Based on your answers to 19 and 20, what is the probability that a lead has oil present given that it tests positive? In other words, what is P(A|B) based on these percentages?
$\frac{5.7\%}{10.4\%} = 0.548$, or 54.8%.

Questions 16 and 17 may be difficult for some students, and the instructor might remind the students that $P(A\&B) = P(B|A)P(A)$ and $P(A\&B^C) = P(B^C|A)P(A)$ to answer question 16, and $P(A^C\&B^C) = P(B^C|A^C)P(A^C)$ and $P(A^C\&B) = P(B|A^C)P(A^C)$ to answer question 17.

The students should notice that their answers to questions 15 through 18 are exactly the same as their answers to questions 8 through 11, divided by 1000. Their final answer to question 21 will be the same as their answer to question 14, but this time, they will use the formula $P(A|B) = \frac{P(A\&B)}{P(B)}$ directly.

IV. Interpret the Results

After calculating the probability that a positive-testing lead has oil present in three different ways, it is time for students to investigate what is happening in a more general way. This will ultimately guide to them "discovering" Bayes' theorem.

Using the simulated data and/or the expected frequencies, point out to students that one reason for the surprisingly low probability that a positive-testing lead has oil present is that the number of "false positive" results is relatively high. This is due to the fact that it is very rare for a lead to have oil present at all. The number of "false negative" results, on the other hand, is quite low.

The discovery of Bayes' theorem begins by pointing out to the students that $P(A|B)$, the probability that a lead contains oil given that it tested positive, was determined using $P(A|B) = \frac{P(A\&B)}{P(B)}$ The initial difficulty of this problem is that neither $P(A\&B)$ nor $P(B)$ are given in the problem statement. $P(A\&B)$ can be determined using $P(A\&B) = P(B|A)P(A)$, since both $P(B|A)$ and $P(A)$ are given in the problem statement.

$P(B)$, the probability that a lead tests positive, is a little more complicated to calculate. However, the students have seen in the activity that a lead can test positive in two ways: either oil is present

and it (correctly) tests positive, or oil is not present and it (incorrectly) tests positive. The probability of the former situation is $P(A\&B)$ again, given by $P(B|A)P(A)$. The probability of the latter, $P(A^C\&B)$, is equal to $P(B|A^C)P(A^C)$. Since these are the only two cases that produce a positive test result, and the two cases are mutually exclusive, the sum of these two probabilities is $P(B)$ itself. That is, $P(B) = P(B|A)P(A) + P(B|A^C)P(A^C)$. This can be demonstrated to students using the concrete examples of questions 18 through 21.

Finally, students have arrived at Bayes' theorem: putting together all of the previous calculations, they can arrive at $P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|A^C)P(A^C)}$. It should be made clear to students that despite the apparent complexity of this formula, it is exactly what they used to calculate their answers in questions 15 through 21. Each piece of this formula has been seen on its own, and students should be able to think of each "chunk", $[P(B|A)P(A)]$ for example, as having a significant meaning within the context of the problem.

Teachers may also want to use ACPPROGS, which is a "solutions" version of the ACPPROG program. This program is included in the ACPPROG program bundle. In addition to the information given by ACPPROG, ACPPROGS will also display the proportion of occurrences of event A when event B occurs, and the proportion of non-occurrences of event A when event B does not occur. Finally, the theoretical probabilities of each category, and the probabilities $P(A|B)$ and $P(A^C|B^C)$ will be displayed.

Assessment

After completing this project, students will have several tools for calculating Bayes' theorem. In the assessment problems below, students can use either expected frequencies or probabilities to solve the problems; students may prefer to use the formal statement of Bayes' theorem, or use the 2x2 charts found in the activity. Instructors may require one particular method to be used, or may allow any of these methods to solve the problems. Instructors may also want to have the students guess at the answers before performing any calculations. Assessment problem 2 requires students to use their understanding of Bayes' theorem to extend the usual formula to account for the three possible factories in which the defective product can be made.

1. Of the emails sent to me, 10% are spam. If an email is spam, my spam filter will correctly identify it as spam 99% of the time. If an email is not spam, my spam filter will incorrectly identify it as spam 3% of the time. Email that is identified as spam is moved to a special spam folder.

a. If I open an email that is identified as spam, what is the probability that it is actually spam?

b. If I open an email that is identified as spam, what is the probability that it is not actually spam?

c. If I open an email that is not identified as spam, what is the probability that it is actually spam?

d. If I open an email that is not identified as spam, what is the probability that it is not actually spam?

e. Compare your answers to problems a-d with the given information that 10% of my email is spam. Is this spam filter effective at keeping spam out of my inbox? Should I occasionally check my spam folder for non-spam emails?

2. A manufacturer uses three different factories to produce its products. Factory A produces 25% of the products, Factory B produces 45% of the products, and Factory C produces 30% of the products. Of the products produced at Factory A, 5% are defective. Of the products produced at Factory B, 3% are defective. Of the products produced at Factory C, 6% are defective.
If a product is chosen at random,
a. …what is the probability that is was made at Factory A?

b. …what is the probability that it was made at Factory B?

c. …what is the probability that it was made at Factory C?

If a product is found to be defective,
d. …what is the probability that is was made at Factory A?

e. …what is the probability that it was made at Factory B?

f. …what is the probability that it was made at Factory C?

g. Compare your answers to problems a-c with those of d-f. How did the probabilities change with the additional information that the product was defective? Do these changes make sense to you?

Answers

1. a. Let S be the event that an email is spam, and F be the event that the filter identifies the email as spam. Then $P(S)$=.1, $P(F|S)$=.99, $P(F|S^C)$=.03. Using Bayes' theorem,

$$P(S|F) = \frac{P(S\&F)}{P(F)} = \frac{P(F|S)P(S)}{P(F|S)P(S) + P(F|S^C)P(S^C)}$$

We also know that $P(S) = .1$ $P(S^C)$=.9, $P(F|S)$=.99, and $P(F|S^C)$=.03. Therefore, $P(S|F) = \frac{(.99)(.1)}{(.99)(.1)+(.03)(.9)} = .7857$

b. Similarly, $P(S^C|F) = \frac{(.03)(.9)}{(.03)(.9)+(.99)(.1)} = .2143$

c. $P(S|F^C) = \frac{(.01)(.1)}{(.01)(.1)+(.97)(.9)} = .0011$

d. $P(S^C|F^C) = \frac{(.97)(.9)}{(.97)(.9)+(.01)(.1)} = .9989$

e. The spam filter is very effective at keeping spam out of the inbox, as less than 1% of the email on my inbox is spam. However, I should occasionally check my spam folder for non-spam emails, as over 20% of emails in the spam folder are not actually spam.

2. Let A be the event that a product was made at Factory A, B the event that it was made at Factory B, and C the event that it was made at Factory C. Let D be the event that the product is defective. Then,

a. $P(A) = .25$

b. $P(B) = .45$

c. $P(C) = .30$,

We know that $P(D|A) = .05$, $P(D|B) = .03$, and $P(D|C) = .06$. Therefore,

d. $P(A|D) = \frac{(.05)(.25)}{(.05)(.25)+(.03)(.45)+(.06)(.30)} = .284$

e. $P(B|D) = \frac{(.03)(.45)}{(.05)(.25)+(.03)(.45)+(.06)(.30)} = .307$

f. $P(C|D) = \frac{(.06)(.30)}{(.05)(.25)+(.03)(.45)+(.06)(.30)} = .409$

g. With the additional information that the product was defective, the probability that the product was made at factory A went up slightly, the probability that it was made at factory B decreased significantly, and the probability that it was made at factory C went up significantly. This makes sense, as factories A and C produce a higher percentage of defective products than factory B. Factory C's probability increased more than that of factory A because it has a higher rate of defective products, and because more products are made at factory C than at factory A.

Possible Extensions

Students can come up with their own questions that involve Bayes' theorem. Any probability that can be influenced by additional information can be turned into a problem involving Bayes' theorem. For example, suppose the students joke that their science and math teachers are conspiring to give quizzes on the same day. This can be turned into a Bayes' theorem problem: If the science class comes before the math class, they can estimate the percentage probability that they have a math quiz given that they had a science quiz earlier in the day. The question can then be asked, "Suppose you know you have a math quiz tomorrow. What is the probability that you will also have a science quiz tomorrow?" Students will need to estimate the relevant numbers to solve this problem.

References

1. Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report, ASA, Franklin et al., ASA, 2007. http://www.amstat.org/education/gaise/

2. Common Core State Standards for Mathematics, Common Core State Standards Initiative (CCSSI). Released June 2, 2010. http://www.corestandards.org/

3. Principles and Standards for School Mathematics. National Council of Teachers of Mathematics (NCTM). Reston, VA: NCTM, 2000.

4. This activity is based on the following:
When 95% Accurate Isn't: Exploring Bayes' Theorem. CadwalladerOlsker, T, Mathematics Teacher, 104(6), p. 426–431, 2011.

When 95% Accurate Isn't Activity Sheet

When exploring an area for oil[1], oil company surveyors will look for *leads*, formations on the Earth's surface that indicate the possibility of oil beneath the surface. However, just because it is *possible* to find oil beneath these leads does not mean that there *is* oil. Suppose that only 6% of leads actually have oil present. The oil company's surveyors will run a series of tests to determine the likelihood of the presence of oil. The surveyors' tests are 95% accurate for both the presence and the absence of oil: If oil is present, the surveyors' tests will reveal that oil is present with a probability of 0.95, and if oil is absent, the surveyors' tests will reach a negative conclusion with a probability of 0.95. If a lead does, in fact, test positive, the company will drill an exploratory well to determine if oil is actually present. Such exploratory wells are expensive to drill, so the oil company wants to know: what is the probability that a positive-testing lead actually has oil present?

In order to explore this question, you will simulate testing 1000 leads using your graphing calculator and the ACPPROG program.

Each lead has two events that will be measured, and each event has two possible outcomes: The event of the presence of oil has possible outcomes of "Oil" or "No Oil"; the event of the surveyors' test has possible outcomes of "Positive" or "Negative". Let Event A represent the event that the lead contains oil, and Event B represent the event that the surveyors' test is positive.

1. What do the following mean, in English?

$P(A)$

$P(B|A)$

$P(B^C|A^C)$

2. What are the values of each of these probabilities, according to the problem?

$P(A) =$

$P(B|A) =$

$P(B^C|A^C) =$

3. If the surveyors' test of a lead is positive, what would you guess to be the probability that the lead actually has oil present?

When we look at both events, A and B, there are four total possible outcomes:

- The lead contains oil, and the test is positive, $(A\&B)$,
- The lead contains oil, but the test is negative, $(A\&B^C)$,
- The lead does not contain oil, but the test is positive, $(A^C\&B)$,
- The lead does not contain oil, and the test is negative, $(A^C\&B^C)$.

[1] Typically, areas are explored for any kind of hydrocarbon deposit, including oil and natural gas, but for the sake of simplicity, we will use the word "oil" to represent both oil and natural gas deposits.

Simulated Data

4. We'll estimate the probability of each possible outcome by simulating tests of a large number of leads. Using the program ACPPROG and your values from question 1, generate 1000 data points. ACPPROG uses the –sign to represent the complement, so –A represents A^C, for example. Record the values generated by ACPPROG here:

$(A\&B)$:
$(A\&B^C)$:
$(A^C\&B)$:
$(A^C\&B^C)$:

Arrange these values into the following table:

	Positive Test	Negative Test
Oil Present		
No Oil Present		

(Note: The sum of all four boxes should be equal to 1000 total leads.)

5. How many leads tested positive for oil?

6. Of those leads that tested positive for oil, how many actually have oil present?

7. Based on your answers to 5 and 6, what is the probability that a lead has oil present given that it tests positive? In other words, what is $P(A|B)$ based on this simulated data?

Expected Frequencies

8. Suppose the oil company tests 1000 leads. We can expect 6% of those leads to have oil. How many leads should we expect to have oil?

9. Of those leads that contain oil, how many should we expect to (correctly) test positive for oil, and how many should we expect to (incorrectly) test negative?

10. Of those leads that do not contain oil, how many should we expect to (incorrectly) test positive for oil, and how many should we expect to (correctly) test negative?

11. Summarize your answers to 9 and 10 in the table below:

	Positive Test	Negative Test
Oil Present		
No Oil Present		

(Note: The sum of all four boxes should be equal to 1000 total leads.)

12. How many leads should we expect to test positive for oil?

13. Of those leads that we expect to test positive for oil, how many should we expect to actually have oil present?

14. Based on your answers to 12 and 13, what is the probability that a lead has oil present given that it tests positive? In other words, what is $P(A|B)$ based on these expected values?

Probabilities

Suppose the oil company tests a large (but unspecified) number of leads.

15. What percentage of those leads to we expect to have oil present?

16. Of those leads that contain oil, we expect 95% to correctly test positive for oil. What percentage of *all* of the leads should we expect to (correctly) test positive, and what percentage of *all* of the leads should we expect to (incorrectly) test negative?

17. Of those leads that do not contain oil, we expect 95% to correctly test negative for oil. What percentage of *all* of the leads should we expect to (incorrectly) test positive, and what percentage of *all* of the leads should we expect to (correctly) test negative?

18. Summarize your answers to 16 and 17 in the table below:

	Positive Test	Negative Test
Oil Present		
No Oil Present		

(Note: The sum of all four boxes should be equal to 100% of the total leads.)

19. What percentage of leads should we expect to test positive for oil?

20. Of those leads that we expect to test positive for oil, what percentage should we expect to actually have oil present?

21. Based on your answers to 19 and 20, what is the probability that a lead has oil present given that it tests positive? In other words, what is $P(A|B)$ based on these percentages?

ACPPROG Program Source Code

Below is the source code for the program ACPPROG. The code below may be modified or re-rewritten for use in other calculators or programs, such as Microsoft Excel. The following should be entered as five separate programs into the calculator, but only ACPPROG will be directly used. The ACPPROG.zip file also includes the programs ACPPROGS, DISPDATC, and DISPPRB. ACPPROGS is a "solutions" version of the program that displays the conditional probabilities $P(A|B)$ and $P(A^C|B^C)$ based on the generated data, as well as the theoretical probabilities calculated by Bayes' Theorem.

```
PROGRAM:ACPPROG
:prgmGETPRB
:prgmCALCPRB
:prgmGENDATA
:prgmDISPDATA
:Pause
:ClrHome

PROGRAM:GETPRB
:ClrHome
:Input "P(A)? ",A
:Input "P(B,A)? ",B
:Input "P(-B,-A)? ",C

PROGRAM:CALCPRB
:(A*B)→K
:(A*(1-B))→L
:((1-A)*(1-C))→M
:((1-A)*C)→N
:(K/(K+M))→Q
:(N/(N+L))→R

PROGRAM:GENDATA
:ClrHome
:Input "NO. OF DATA? ",I
:0→S
:0→T
:0→U
:0→V
:0→X
:For(X,1,I)
:rand→Y
:If Y<K:Then
:S+1→S:Else
:If Y<(K+L):Then
:T+1→T:Else
:If Y<(K+L+M):Then
:U+1→U:Else
```

```
:V+1→V
:End:End:End:End

PROGRAM:DISPDATA
:ClrHome
:Output(1,1,"GENERATED DATA:")
:Output(2,1,"A+B:")
:Output(2,10,S)
:Output(3,1,"A+-B:")
:Output(3,10,T)
:Output(4,1,"-A+B:")
:Output(4,10,U)
:Output(5,1,"-A+-B")
:Output(5,10,V)
```