EGN 1932 Homework 4 (HIV Testing) 20 Points

Consider a medical condition (HIV infection, say, as a concrete example) for which there is a diagnostic test. Any such test can make two types of error: (1) the FALSE POSITIVE rate is the fraction of uninfected individuals for whom the test incorrectly produces a positive result, and (2) the FALSE NEGATIVE rate is the fraction of infected individuals for whom the test incorrectly produces a negative result.

Suppose that the false positive rate is 2% and the false negative rate is 1%; and suppose that the PREVALENCE (the fraction of the population that is infected) is 0.1% (I concocted these numbers for simplicity, but I think that they are reasonable.)

The following BASIC code simulates the HIV test. (p is the prevalence, f1 is the false positive rate, and f2 is the false negative rate; (n2 + n3)/1000000 is the overall accuracy of the test, and n2/(n1 + n2) is the fraction of those individuals who test positive and who are, in fact, infected with HIV.)

```
RANDOMIZE
100 INPUT p,f1,f2
110 FOR i = 1 TO 1000000
120 IF RND > p THEN 150
130 IF RND > f2 THEN n2 = n2 + 1
140 GOTO 160
150 IF RND < f1 THEN n1 = n1 + 1 ELSE n3 = n3 + 1
160 NEXT i
170 PRINT (n2 + n3)/1000000, n2/(n1 + n2)
```

Before you run the code, estimate (a) the probability that the test result is accurate, and (b) the probability that a person who tests positive is, in fact, infected with HIV. Write your estimates here:

(a)

(b)

Run the code and compare the simulation results with your predictions (above). Do they agree? If not, can you explain why?