# 06N:216 *Data and Decisions*

**Homework 2 Partial Answers**

1. Do out the “Bayes’ Rule” calculations needed to reproduce the conclusion of the article “Propensity to Abuse-Propensity to Murder?” from *Chance* magazine. Use the following events:

M: Murdered woman was killed by current or former partner,

E: Murdered woman was killed by someone else,

## A: Murdered woman has a known history of abuse by current or former partner,

N: Murdered woman has no history of abuse by current or former partner.

The goal is to compute P(M|A). Do this under the assumption that P(M) = 0.29, P(A|M) = 0.5, and P(A|E) = 0.05, as suggested in the article. The article states that under the above assumptions, the “posterior odds ratio” is 4:1, which is the same thing as saying that P(M|A) = 0.8, or 80%. *Ignore the “odds form of Bayes’ rule” given in the article. Hint: consider a population of 1,000,000 men whose wives were murdered.*

**Consider a population of 1,000,000 men whose wives were murdered. Of the associated unfortunate spouses, 29% were murdered by their husbands and 71% by someone else----yielding numbers of 290,000 and 710,000 respectively. Of the 290,000, 50% or 145,000 were abused by their spouses. Of the 710,000, 5% OR 35,500 were abused by their spouses. This gives us a total of 145,000 + 35,500 = 180,500 who were abused by their spouses. Of this 180,500 who were abused, 145,000 were murdered by their husband. Thus the probability that the husband did the murder GIVEN a history of prior abuse is:**

**145,000/ 180,500 = 80.33%.**

**See also the EXCEL spreadsheet for an equivalent set of calculations using the conditional and joint probability notations.**

1. As a marketing manager for a fast food company, you are developing menu item for the Asian market called Fugu McNuggets. Based off historical data on past ventures, new products introduced in the marketplace result in a “high” sales category 10% of the time and a “low” sales category 90% of the time. You are considering a test marketing campaign to better gauge the sales of this new product. It is estimated that a test market has the following accuracies: for high sales products, consumer test reaction is positive 70% of the time, inconclusive 25% of the time, and negative the remaining percent of the time; for low sales products, then consumer test reaction is negative 55% of the time, inconclusive 30% of the time, and positive the remaining percent of the time. Given a positive test market outcome, what is the probability of high sales and the probability of low sales? Given a negative test market outcome, what is the probability of high sales and the probability of low sales? Given an inconclusive test market outcome, what is the probability of high sales and the probability of low sales? Hint: consider a population of 100,000 new products.

**Let us imagine a population of 100,000 new products. About 10%, or 10,000 could be expected to be high sales category products and 90,000 could be expected to be low sales category products.**

**If it is a high sales product, the test market can be expected to be positive 70% of the time, so 7000 out of the 10,000 high sales products would be expected to get a positive test. 25% or 2500 would be expected to get an inconclusive test result and the remaining 5% or 500 would be expected to get a negative test result.**

**If it is a low sales product, the test market can be expected to be positive 15% of the time, so 15% of the 90,0000 = 13,500 low sales products would be expected to get positive test results. Similarly we could expect to see 27,000 (30% of 90,000) inconclusive test results and 49,500 (55% of 90,000) negative test results.**

**Positive Market Test**

**We can expect to see 7000 + 13,500 = 20,500 positive test results, of which 7000 are high sales products. The probability of high sales given a positive market test is 7000/20,500 = 34% (approximately) The probability of low sales given a positive market test is 66% (approximately).**

**Inconclusive Market Test**

**We can expect to see 2500 + 27,000 = 29,500 inconclusive test results. The probability of high sales given an inconclusive test is 2500/29500 = 8.5% (approximately). The probability of low sales given an inconclusive test is 91.5%.**

**Negative Market Test**

**We can expect to see 500 + 49,500 = 50,000 negative test results. The probability of high sales given a negative market test is 500/50,000 = 1%. The probability of low sales given a negative market test is 99%.**

For the next two problems, assume that the probabilities of having either male or female children are equal---a 50-50 proposition.

1. Suppose you meet a man and learn that he has exactly two children. Suppose further, that you learn that the older child is a boy. What is the probability that both children are boys? Hint: count out the possibilities and see which can be eliminated.

**Without any knowledge at all, here are the possible patterns of children (all equally likely), where the first letter denotes the gender of the firstborn child and the second letter, the gender of the secondborn.**

**BB**

**BG**

**GB**

**GG**

**Since the firstborn is a boy, that means that the GB and GG patterns are impossible, only the first two remain. Of the two remaining possibilities, one and only one (BB) results in two boys. Since both possibilities are equally likely, the probability that both children are boys given that the firstborn is a boy is 50% or ½.**

1. Suppose you meet another man and learn that he has exactly two children. Suppose further, that you learn that at least one of the two children is a boy. What is the probability that both children are boys?

**Now only GG is eliminated, so 1/3.**

1. Sickle cell anemia is an inherited blood disorder characterized primarily by chronic anemia and periodic episodes of pain. The underlying problem involves hemoglobin, a component of red blood cells. Hemoglobin molecules in each red blood cell carry oxygen from the lungs to body organs and tissues and bring carbon dioxide back to the lungs. ----------  
     
   1. If a husband and wife both carry the sickle cell trait, what is the probability that a child born of their union
      1. will develop sickle cell disease,

**Both are coded SA, so the possible patterns in a child are:**

**SS**

**SA**

**AS**

**AA**

**All patterns are equally likely, so the probability is ¼.**

* + 1. will carry the sickle cell trait but not develop the disease, and

**From above, the probability of SA (either AS or SA) is ½.**

* + 1. will be completely free of sickle cell genes?

**From above, the probability of AA is ¼.**

* 1. Suppose that the proportion of males as well as females in a certain population that have the sickle cell trait is 1 out of 12. Suppose further that marriages between men and women occur randomly without respect to sickle cell trait. What proportion of children born in this population will develop sickle cell disease? To keep things simple, assume that people who develop sickle cell disease do not marry.

**The probability of a child having both parents coded as SA is 1/12 x 1/12 = 1/144.**

**The probability that a child born of such a union will have sickle cell disease is ¼. So the overall probability is ¼ x 1/144 = 1/576 = 0.001736.**

1. A particular hypothetical human disease occurs with a probability of 0.1 in males and with a probability of 0.4 in females.

* 1. Assuming that the frequency of males is 0.5 and females 0.5 in a very large population, what is the probability that an individual selected at random from this population will have the disease?

**Let’s look at the possible combinations: (where D means has the disease and M stands for male)**

**MD**

**MN**

**FD**

**FN**

**Now these events are not all equally likely, but they are MECE.**

**We are looking for (note that the intersection of MD and FD is empty)**

**P{MD U FD} = P{MD} + P{FD} =**

**P{MD} = P{Male} x P{D|Male} = ½ x 0.1 = 0.05**

**P{FD} = P{Female} x P{D | Female} = ½ x 0.4 = 0.2**

**So 0.05 + 0.2 = 0.25 is the answer.**

**Can also use a “scenario” approach:**

**Imagine selecting 100,000 people. We would expect 50,000 to be males and 50,000 to be females. Of the males, we would expect 10% or 5000 to have the disease. Of the females, we would expect 40% or 20,000 to have the disease. In total, we would expect 25,000 of the 100,000 to have the disease. So the probability that a randomly selected person would have the disease is 25000/100000 = 25%.**

* 1. Under the same assumptions, what is the probability that an individual will be male and have the disease? Be female and not have the disease?

**Let’s look at the possible combinations: (where D means has the disease and M stands for male)**

**MD**

**MN**

**FD**

**FN**

**Now these events are not all equally likely.**

**We are looking for the probability MD as well as FN.**

**First let’s work out P{MD}:**

**P{MD} = P{Male} x P{D|Male} = ½ x 0.1 = 0.05**

**Scenario approach: Of the 100,000 people we would expect 5000 to be males and have the disease (see above). So the probability that a single person drawn from the population is male and has the disease is 5000/100000 = 5%.**

**Now, let’s work out P{FN}:**

**P{FN} = P{Female} x P{N | Female} = ½ x 0.6 = 0.3**

**Scenario approach: Of the 100,000 people we would expect 50,000 to be female. Of these, 40% would have the disease and 60% would not. So we would expect to see 0.6 x 50,000 = 30,000 females (out of the 100,000 people chosen at random) to be disease free. The probability of a randomly chosen person being female AND disease free is, therefore, 30,000/100,000 = 0.3**