MSM 709
Solutions to selected problems

Ex.1. Let M = person took math, B = person took biology. Then $P(M) = \frac{250}{325}$. $P(B) = \frac{197}{325}$ and also $P(M \text{ and } B) = \frac{150}{325}$.

1. $P(M \text{ or } B) = \frac{250}{325} + \frac{197}{325} - \frac{150}{325} = \frac{197}{325} = 0.914$
2. $1 - 0.914 = .086$
3. $P(M \text{ and not } B) = \frac{250}{325} - \frac{150}{325} = \frac{100}{325} = 0.308$
4. $P(B \mid M) = \frac{150}{250} = 0.6$

Ex.2. 2. No, H and R are not independent since $P(H) \neq P(H \mid R)$
Also, H and R are not disjoint (i.e. not mutually exclusive) since they have outcomes in common. In fact, H is a subset of R.

3. No, R and S are not independent, since $P(R) \neq P(R \mid S)$. R and S *are* disjoint, since there are no cards that are both red and a spade.

4. No, H and S are not independent, since $P(H) \neq P(H \mid S)$. H and S *are* disjoint since there are no cards that are both a heart and a spade.

Ex. 3 Let M = male and H = graduated high school. We are given $P(M) = 0.6$, $P(H) = 0.8$ and $P(M \text{ or } H) = 0.95$. To solve the given problems, first we must find the probabilty of the intersection of the two events, $P(M \text{ and } H)$. Using the addition rule of probabilty, we get

$$0.95 = 0.6 + 0.8 - P(M \text{ and } H) \text{ so } P(M \text{ and } H) = 0.45$$

Then $P(M \mid H) = \frac{0.45}{0.80} = 0.5625$

Also, $P(H \mid M) = \frac{0.45}{0.60} = 0.75$

$P(H \mid M^C) = \frac{0.35}{0.40} = 0.875$

Ex. 4 $0.72 \times 0.27 = 0.1944$

(There was no Ex. 5)

Ex. 6 Let D = the person has the disease and let Pos = the test came up positive. I'll also use Neg = the test came up negative (rather then the complement notation for this event.)

The information given to you is that the test is 95% accurate, or it's correct 95% of the time. This means that if a person has the disease, the prob. the test is positive is 0.95. Another way of thinking about it is that 95% of all those with the disease would have a positive test (and so 5% of those with the disease would have a negative test.). We can express this in terms of conditional probability as $P(Pos \mid D) = 0.95$ and $P(Neg \mid D) = 0.95$. The accuracy statement also means that 95% of all those *without* the disease would have a negative test (and 5% of those with out the disease would have a positive test.) We can write

$P(Neg \mid D^C) = 0.95$ and $P(Pos \mid D^C) = 0.05$

We are also told that 2% of the population has the disease, or $P(D) = 0.02$. So that means 98% of the population does not have the disease: $P(D^C) = 0.98$.

We are asked to find the probability of a false positive. That is, we are asked to find the probability that if a test is positive, the person does not have the disease after all. This is $P(D^C \mid Pos)$. Another way of thinking about this question is to ask, out of all the positive tests, what proportion came from people who don't actually have the disease? Asked this second way, it is clear that our first step is to find the overall probability that a test will be positive.

$P(Pos) = P\left(D \text{ and } pos\right) + P\left(D^C \text{ and } pos\right)$. That is, we have to add up the positive tests from those with the disease to the positive tests from those without the disease.
So, continuing: $P(Pos) = P\left(D \text{ and } pos\right) + P\left(D^C \text{ and } pos\right) = 0.02 * 0.95 + 0.98 * 0.05 = 0.068$ (i.e.

95% of the 2% with the disease are positive tests, and add to that the 5% of the 98% without the disease whose tests are positive.)

Now, we still are trying to find $P(D^C \mid Pos) = \dfrac{P\left(D^C \text{ and } pos\right)}{P(Pos)}$ by the definition of conditional probability.
The numerator is $0.98 * 0.05 = 0.049$, and we already found that $P(Pos) = 0.068$. So, the answer to the problem is $P\left(\text{false positive}\right) = P(D^C \mid Pos) = \dfrac{0.049}{0.068} = 0.72$.
So, even though the test is 95% accurate there is still a 72% chance of a false positive test!

What is going on here? Well, the main issue is that the overwhelming majority of the population (98%) does NOT have the disease, but the 5% of incorrectly positive tests from that portion of the group contributes the larger portion of all the positive tests. If you were testing for a condition that occured in about 50% of the population , you wouldn't see such a large discrepency between accuracy of the test and the chance of a false positive.