PROBLEM SET 1

1. (a) For this study, I would say a randomised experiment would be more suitable because having a group that undergoes the treatment and another that does not will help very clearly prove that the vaccine works if the majority of the treatment group does not fall sick while the majority of the control group does fall sick.

   (b) I would say an observational study would be more suitable for this because you cannot do an intervention here. We want to study how people are on a particular social topic. So ideally we want to see how people really are in their day to day lives. Are they reacting negatively or positively is something that can be noted only by observing how they are in their daily life.

   (c) A randomised experiment would be effective to test if laptops in classrooms benefit or hurt exam scores. Because if either of the groups shows statistically significant differences in their performance, then it will be very evident and allows a clear conclusion about the impact of laptops in classrooms.

   (d) An observational study might be suitable here because there is no treatment to be done here. Just observing the people with income levels over and below 100k as to which parties they voted for will be sufficient to see if there really is a pattern here.

   (e) A randomised experiment will be suitable for this study. Because eating bacon or not eating bacon would be an intervention that can be used to carefully assess with a base group (control group) to see if bacon really causes colorectal cancer or not with certainty.

2. (a) This would be an observational study as in a blind randomised experiment you don't adjust any of the factors. People are assigned randomly to groups without checking anything about them.

   (b) We need to be careful here as an observational study can be influenced by unforeseen factors like food and fitness. Because it could be the case that using contraceptive pills makes one more hungry and lethargic and as result they eat any kind of food which could be very unhealthy they can get their hands on at the moment and avoid exercising because of how their body feels while in reality the pill in itself might not be causing cancer.

   (c) Some other factors that could vary in general between women are family history, diet, sleep patterns, fitness, the kind of environment and elements they are exposed to on a daily basis, and mental health conditions.

3. (a) The probability of this being a biassed study is very high. If the respondents are all people sitting behind a computer screen, then, for all we know, 10 of those respondents could actually be just 1 person using 10 different accounts online and is being paid to

just answer whatever on a set of multiple choice questions. Additionally, the given information about this survey doesn't state whether the respondents are all from the U.S. or not. They could be from any part of the world as well. In general, on the internet, there can be a lot of manipulation being done in favour of ulterior motives. While in reality, the public's response could be vastly different. Moreover, some might even be giving random answers for the sake of being paid for it because they have no reason to care for the integrity of the survey. While it's true that even the general public could be biassed, the sample of respondents will be more random and representative of the general public. So this survey definitely does not give an unbiased picture of how the U.S. population would respond.

As for the direction of the bias here about the outcome being an overestimation, it depends. The survey could be overestimating either way because with the given information, there are other factors that the outcome's bias will be dependent on. For example, the majority of respondents could be people from countries like China and India. Depending on how the culture around social issues in these countries are, the outcome can largely shift towards one side. Another factor could be how people between the ages of 18 - 44 differ in opinions compared to people aged 45 and above. Since the respondents majorly consist of people in the age group of 18-44, the survey results will be biassed in their favour. As a result, the given information is not sufficient to conclusively identify the direction of bias.

(b) If I were to carry out research on this, I would first draft the survey form. We will also need to know the respondents' age, education, whether they're a citizen of the U.S, which state they're from originally and so on. Then I would form a team. After which, I would divide the team into subteams and assign them to different regions of America. And then I would get them to go to universities, companies, and high density population areas and get them to take the survey in person. If each subteam surveys 100 people in a day and there are about 10 subteams, we would already have 1000 respondents in just one day. At this rate, our sample can easily be of big size. Depending on how much time the budget will allow us to do this survey, we could survey for as long as we can and get a reasonable sample size consisting of people from different areas of America. This will give us a good idea about how the general U.S. adult public would respond to our statement. Finally, I would compile all of this data and then analyse it to derive the answer to our research question.

4. (a) It will be 1/52 as there is only one ace of hearts in a deck of 52 cards.

   (b) There is only one ace of hearts and one queen of spades in a deck. So the probability of getting an ace of hearts on the first card will be 1/52. However, now the deck will have 51 cards in total. Which means the probability of getting a queen of spades will be 1/51. If we multiply these two, we get: 1/52 * 1/51 = 1/2652 = 0.0003.

(c) In the previous question the order mattered. However, in this one, since the order does not matter, there are two possibilities: first ace of hearts & queen of spades, second queen of spades & ace of hearts. Hence, the probability will be: 1/52 * 1/51 + 1/52 * 1/51 = 2/2652 = 0.0007.

(d) For this question, it will be similar to the previous one, except, we are replacing the first card and then reshuffling. So the total number of cards when we draw the second one, will be 52. Therefore, the probability will be: 1/52 * 1/52 + 1/52 * 1/52 = 1/1352

(e) There can be two scenarios here. Scenario one is the first card drawn will be an ace of hearts. Then the probability of drawing an ace of hearts again will be 0 as there is only one such card. Scenario two is where a card other than ace of hearts will be drawn. The probability of this should be 51/52. And now the ace of hearts being drawn will be 1/51. These scenario is a dependent event and so we have to consider it this way: 51/52 * 1/51 = 1/52. Now the total probability will be 0 + 1/52 = 1/52.

5. (a) The probability of getting an even number on a single dice is 3/6 because there are 3 even numbers of the total 6 numbers. If we want an even number on all four dice, then the probability will be: 3/6 * 3/6 * 3/6 * 3/6 = 1/16.

   (b) For this, if we just find the complement of opposite outcome, then we get: 1 - P(No dice show six) = 1 - (⅚ * ⅚ * ⅚ * ⅚) = 1 - (625/1296) = 671/1296.

   (c) If we were to calculate the probability of the sum being > or = 6, then we would have to individually calculate the probabilities of when the sum is = 6,7,8 and so on. Instead, taking the complement of probability of the sum being < 6 would be easier as there are only two possibilities 4 and 5. For 4, there is 1,1,1,1. For 5, there is 1,1,1,2; 1,1,2,1; 1,2,1,1; 2,1,1,1. Adding these probabilities and subtracting the answer with 1 will give us the desired result. The answer would be 1 - (1/1296 + 4/1296) = 1 - 5/1296 = 1291/1296.

6. (a) We need to find the number of combinations 5 gimmels and 5 hehs have and multiply that with the probability of 1 case (HHHHHGGGGG). This will give us the probability of any combination of five gimmels and five hehs. The total number of combinations will be 10!/(5! * 5!). The probability of 1 case will be (¼)^10. Now we multiply the two = 10!/(5! * 5!) * (¼)^10 = (10 * 9 * 8 * 7* 6)/5! * 1/1048576 = 63/262144 = 0.0002.

   (b) The probability of 1 spin not being a nun or shin is 2/4. For 10 spins it will be (2/4)^10 = 1/1024.