Date: __/__/____

S520 - INTRODUCTION TO STATISTICS
Problem Set 1

Ans. 1(a) This Question is better answerd using a randomised experiment. Because it's a cause & effect relation where we're trying to determine if flu vaccine causes no flu. For cause & effect studies, we generally use randomised controlled experiment

(b) This question is better answered using an observational study. Here, we just have to observe a sample from the population. We carry out an observational study when we cannot control the individuals that are part of the study.

(c) This question is better answered using a randomised experiment. Because it's a cause & effect relation where we're trying to prove that no laptops = better exam scores in class. The experiment can be done by taking students taking the same class & of relatively similar IQ for 2 different years.

(d) Observational Study. Here we cannot control the individuals that are part of the study. So we'll just have them answer the survey & do an observational study on them

(e) Randomised Experiment. It's a straight forward cause & effect study which can have 2 cohorts of people. One that eats bacon & the

Date: __/__/____

other cohort can be given a placebo that looks like bacon to let the experiment be blind

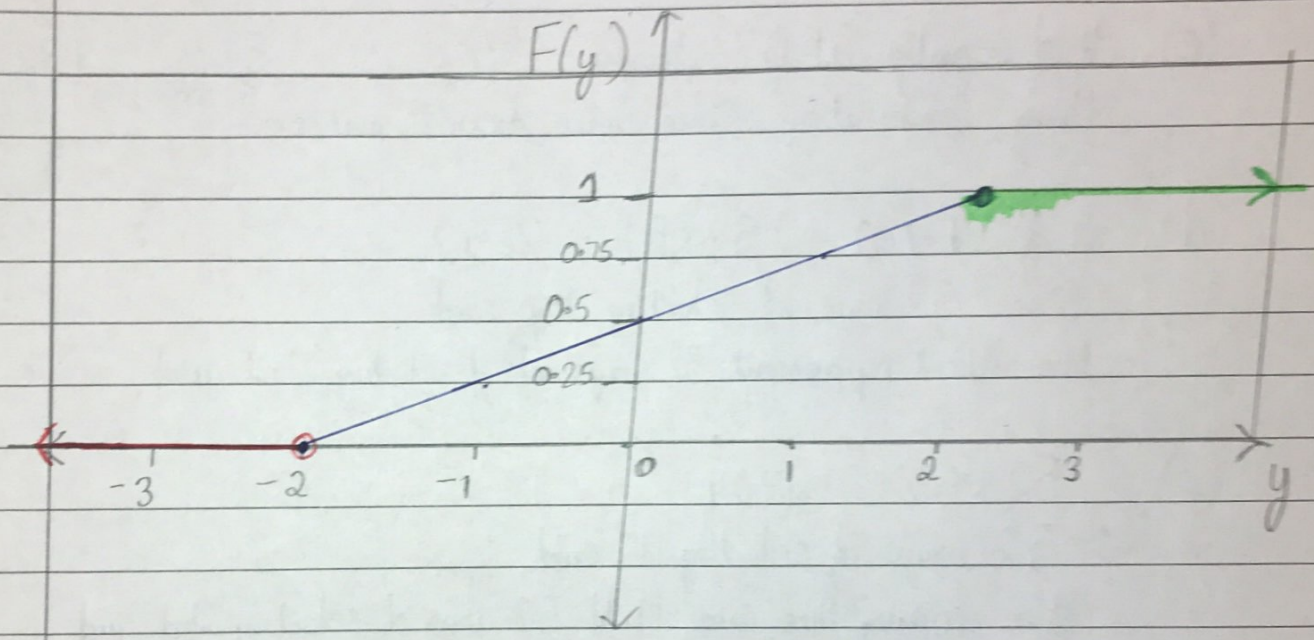
Ans. 2(a) This Study was an observational study since it had no control over the individuals who were part of the study (women). Also, there is no mention of any placebo pill given to women who do not eat the oral contraceptives in the name of "blind" randomised experiment

(b) No, this doesn't prove that the pill causes cervical cancer. Because, the study was done for only a small population of women in California & hence we cannot extrapolate this conclusion for the entire population of women. Also there could be more confounding variables that are not accounted for to prove the causality other than age, educat" & marital status

(c) There could be other confounding variables / factors in the study that relate use of oral contraceptives to cervical cancer in women other than age, education & marital status like ethnicity where women of different ethnicities (Asian, Whites, Hispanic) have different susceptibilities to cervical cancer. Another factor could be frequency of sexual encounters which is higher for women using oral contraceptives & hence, higher risk of cervical cancer.

Date: __/__/____

Ans 4(a) $F(y) = \begin{cases} 0, & y < -2 \\ \frac{y+2}{4}, & -2 \le y < 2 \\ 1, & y \ge 2 \end{cases}$



(b) $F(y) = \begin{cases} 0, & y < 1 \\ \frac{y}{5}, & 1 \le y < 4 \\ 1, & y \ge 4 \end{cases}$

Ans 5(a) $\boxed{{}^{15}C_6 = 5005}$ # ~~Seq~~ tosses = 15

→ # ordered sequences out of 2^{15} with exactly 6 heads

(b) $\boxed{{}^{10}C_2 \times {}^{5}C_4 = 225}$ both are independent events

Hence using multiplication rule

#learnthesmarterway

Date: __/__/____

Ans. 6 (a) # of equally likely outcomes = 52
Since 52 ways of picking a card

(b) # of equally likely outcomes = $^{52}P_2 = \frac{52!}{50!} = 2652$
Using permutations since order matters

(c) # of equally likely outcomes = $^{52}C_2 = \frac{52!}{50! \times 2!} = 1326$
Using Combinations since order doesn't matter

(d) # of E.L.O = 52 x 51 = 2652
Since 52 ways of selecting top card
Now without replacement, 51 ways left of selecting 2nd card

(e) 52 x 52 = 2704
52 ways of selecting 1st card
After replacing this card, still 52 ways of selecting 2nd card

Ans. 3 (a) No. This doesn't seem like a statistically unbiased answer to the question. There appears to be 2 biases in this study.
(i) Out of the people who have an opinion, most (>70%) agree that sports teams should take public stance on social causes. But out of the total people surveyed, >50% are democrats & just 22% are republicans. Democrats, in general have a liberalist opinion about social issues & hence this survey's results seemed biased in favour of democrats
(ii) There also appears to be a gender bias ∵ there is no mention about the no. of men & women being surveyed.

#learnthesmarterway

Date: __/__/____

(b) Given the required resources to perform this study, I would tweak the approach a little bit

(i) First of all, get more people for the survey. This is the most important part as we need the sample size to be big to make reliable inferences. So, I'd have multiple trusted online survey platforms carry out the surveys.

(ii) The sample for the population should represent the population. So, we need to improve for statistical biases in the previous study like:

(A) Gender Bias: There needs to be an equal representat^{n} of both genders & also include transgenders in this survey to get a better picture of the sample

(B) Political inclination: The sample should include both democrats (liberalist view) & Republicans (conventional view) & both of them should have similar representat^{n}.

(iii) If bias for certain variables cannot be adjusted due to lack of data, then we need to do weighting or assigning weights to variables to prevent bias