STA 4010 HW3

潘力硕 115010208

T1

(a):

I think the breast cancer dearth rate(per 1000 women) in the group of people in Treatment and Screened shows the efficacy of treatment.

Reason: because, our problem is: "Do screening programs speed up detection by enough to matter?" So in the treatment group, the target people should be the people who have accepted the screen treatment.

(b):

we compare the death rate of the Treatment and Screened group with the control group, there is no significant difference. Also we compare the death rate of the total Treatment group and control group, there is no significant difference.

So we draw the conclusion that screen program can not speed up the detection of other cancers.

(c):

I think this is not a good idea, because doing this randomized trails(not randomized strictly speaking) we want to generalize the result to apply to all the patient not only women because our aim is to answer the question that "Do screening programs speed up detection by enough to matter?". So in the subjects, we should choose all the possible subject which may happen in the real situation.

T2

I totally agree with the statement that "Association is not causation". Specificlly, I think "causation" is a much stronger relationship than "association". "causation" inherently means "association", however, "association" does not necessarily means "causation". For example, given the dataset, if we do the regression and find out that one of the feature is highly related to the outcome. Let's say "study hours" is highly related to the "GPA". In this case, we can only draw the conclusion that "study hours" is positively "associate" with the "GPA", we can not draw the conclusion that high "study hours" "cause" the high "GPA". Because, we can see the data points that do not have high "study hour" but with high "IQ" or "study method" also get high "GPA". So, this also fit our intuition that some students who is smart or has good study method also get high GPA although they do not spend much time on the subject. This is one perspective that support this statement.

For another perspective, we observe that high "GPA" "associate" with the high "study hour". We could also interpretate it in the way that high "GPA" makes students tend to be more involved into the study which means higher "study hours". Then the relationship between the feature and outcome becomes that the outcome "cause" the feature which is the opposite of the statement "higher study hours cause the higher GPA".

Above all, we can draw the conclusion that the statement "Association is not causation" is correct.