

# STAT616 - HW #1

*Elliot Smith*

*1/25/2018*

## Problem 1

The answer is (ii) Drinking is associated with smoking, and alcohol causes liver cancer. The premise behind a confounding variable is, in this scenario, that drinkers tend to smoke more, and as a result, while the actual cause of liver cancer is alcohol, the act of smoking is being inappropriately labeled as the cause of liver cancer. In summary, while alcohol is the cause of liver cancer, because so many drinkers also smoke, smoking is being attributed as the cause of the liver cancer, when the alcohol is the true culprit.

## Problem 2

I believe that this can be explained by the fact that the design of the experiment was relayed to the participants. The question wording "... sometimes treated with zinc sulfate" leads me to believe that the solution is not always successful at treating the malady, otherwise it would always be used. I will identify the following groups:

- Group 1 — 1st Half - Placebo / 2nd Half - Placebo
- Group 2 — 1st Half - Placebo / 2nd Half - Zinc
- Group 3 — 1st Half - Zinc / 2nd Half - Placebo
- Group 4 — 1st Half - Zinc / 2nd Half - Zinc

Assuming that all group sizes are equal, if I am a participant in the experiment and I am given the Placebo in the first half of the study, then I know there is approximately a 50% chance that I will be given Zinc for the second half of the study. I will now further break the groups down:

- Group A — 1st Half - given Zinc Sulfate and the treatment was successful
- Group B — 1st Half - given Zinc Sulfate and the treatment was unsuccessful
- Group C — 1st Half - given the Placebo

Now, before the first half of the experiment, the participants have no information on whether they are in any of the four original groups. However, after the first half has concluded, the participants will now fall into one of the three new groups. A new element is inserted, the patients in Group A will be fairly certain that they were treated successfully, and as such, will most certainly be able to tell whether or not they are in Group 3 or Group 4 by if their symptoms return (Group 3) or if they symptoms remain treated (Group 4). However, those in Group B and Group C, which make up more than half of the total participants, having seen no improvement in symptoms will infer that they received a Placebo in the first half of the experiment. At this point, those who received a Placebo in the first half of the experiment have a 50% chance of receiving the Zinc Sulfate treatment in the second half. However, since Group B and Group C combined are greater than the total of those who received a Placebo in the first half, a greater number of patients than should believe they will be receiving the treatment. The Placebo Effect kicks in and more patients than should believe that they are now on the Zinc Treatment and will thus "feel better" when its actually the result of perception and not the treatment of Zinc Sulfate.

## Problem 3

### Problem a

I would choose to test the new participants at the beginning of the second year. The reason I would choose to test them during registration (as opposed to not testing them) is because the goal of the fitness program is to measure the effects it has on the participants. The assumption that the investigators are making (by questioning the necessity of testing at the beginning of the year) is that the group that will sign up for the second year will generally be the same as the group that signed up for the first year; we cannot make this assertion. For example, the first year group that signed up may have all been in poor, physical shape and thus were eager to sign up. However, those that signed up for the second year, hearing of the great improvements made by the previous cohort, may be comprised of individuals in even worse shape than the previous year. In this scenario, while the second year cohort may make the same incremental gains in their physical fitness, it would appear as though they did not perform as well as the previous year because we never measured where their level of fitness was beforehand.

### Problem b

The method of analysis that I would like to perform is a 2-sample t-test. By collecting the physical fitness levels of the registrants at the beginning of the trial and at its conclusion, I would ascertain the mean fitness level from each sample and perform a t-test to conclude whether or not the difference in means is significant. This would allow me to conclude whether or not the physical fitness program has a positive effect on those that partake in it.

## Problem 4

### Problem a

It looks like the the groups are not homogeneous aside from the medication based on the fraction of each group that adhered to the treatment protocol. While only 53% of the Nicotinic Acid group adhered to the treatment regiment, 67% of the Placebo group adhered to the treatment; this sizeable difference implies that there is a strong probability that there are other differences between the groups aside from the treatment. For example, if a larger fraction of a group adheres to a treatment, this may imply that they are more concerned about taking care of themselves, and as a result, may have a decreased chance of death than those that are less concerned about their health.

### Problem b

In order to determine the efficacy of the Nitocinic Acid, I would look at the percent of deaths from those the adhered to the treatment from the protocol from the Nicotinic Group (13%) and the Placebo Group (15%). Once we specify that we are only looking at those that adhered, we are controlling for those that adhere as well as the other factors already considered. I do not believe that a valid comparison can be made at this point. Before concluding anything, I would want to run a 2-sample t-test to confirm whether or not the difference in means from the two samples (Nicotinic Group - 13% and Placebo Group - 15%) is significant enough for us to conlude whether or not the treatment is effective at reducing the death rate.