Midterm

We have reached the middle of the term. So, let's do a midterm and then ride off into the sunset of spring break 2023.

Extremely short answer section

(a) Running multiple experiments examining the same relationship between an independent and dependent variable, using variations in the treatment, outcome measure, sample, etc. would be done to increase what?
(b) If the measure the researchers use to capture the dependent variable has little to do with the dependent variable, we would say it lacks what?
(c) Internal validity in experiments is achieved through what means? By doing what?
(d) In this class, the concepts of reverse causality and _____ are frequently mentioned together.
(e) (T/F) If I use a non-random sample in my study I cannot make causal inferences.
(f) If I have **four** people in a group and I want to know the probability/proportion that successfully complete a task, which of the following six numbers are *possible* results?

(a) 0.25, (b) 1.25, (c) 0.5, (d) 0.9, (e) -0.25, (f) 1

Pretty short answer section

(a) How would you (correctly) explain the use/importance/role of a control group?
(b) What would be an example of reverse causality?[1]
(c) Where can experiments take place?
(d) What determines the number of groups in an experiment?

[1] This question is *not* asking you to remember an example from class (though that would be an acceptable answer). It's asking you to remember the concept and describe a scenario where that would occur.

Another Twitter Screenshot?

Below is the result of a Twitter poll that I have no idea why the algorithm wanted me to see.

Just to be sure we're on the same page, this says, for example, that 57.2% of respondents *do* know **and** are male. Now I'm going to ask some probability questions. You can write probabilities either as decimals or fractions.

a) P(male) =
b) P(no) =
c) P(male | no) =
d) Is there a correlation between one's gender and dress/skirt differentiation ability? Why or why not?

I was really hoping there wasn't going to be anything about potential outcomes on this test...

On January 17$^{\text{th}}$, 2023, journalist, book author, and podcaster Jesse Singal published what turned out to tendentious op-ed entitled, "What if Diversity Trainings Are Doing More Harm Than Good?". The New York Times subsequently allowed readers to have their letters to the editor published in a stand-alone article. One of these responses contained the following paragraph,

> During my career in corporate middle management, I was subjected annually to all kinds of training modules: diversity, sexual harassment, anti-money-laundering. By and large those who were receptive to such repetitive training needed it the least. Conversely, those who needed it the most weren't receptive beyond paying lip service.

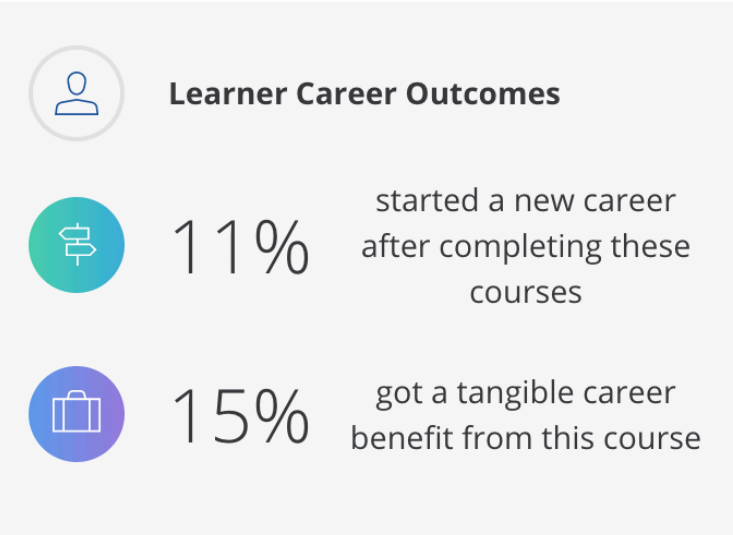
How would you express this using the logic of potential outcomes and treatment effects? Note: your answer can (and should) be a mix of normal, English-language analysis and a symbol or two. There's no one correct answer, but yours should include some idea of independent and dependent variables. A big hint is that you'll want to use `1` and `0` to indicate … things. Maybe multiple different things.

Correlation?

Imagine you read a study reporting that 80% of Silicon Valley CEOs are taking ashwagandha.[2] Evaluate the two claims in the following dialogue.

- "Man, did you see that study on ashwagandha? I need to start taking it if I want that CEO brain."
- "Maybe, bro. But that *is* a spooky high correlation between taking ashwagandha and being CEO"

Truth in advertising



This is an informational advertisement from Coursera.com, a website that hosts university courses that motivated "self-learners" can take.

(a) Does the first number imply a *relationship* between taking the courses referred to and starting a new career? If not, why not?
(b) How would you express as a conditional probability the idea that if you take the courses you get a tangible career benefit?[3]
(c) How would you evaluate the claim that '15% of learners got a tangible career benefit from this course' from potential outcomes perspective? Think about what we know and what we don't know.

[2] An evergreen shrub that grows in Asia and Africa, commonly used for stress and marketed as an "adaptogen."
[3] Don't think too hard about this question – it's just asking you to notate something correctly.

Over/under

Let's think about over-estimates and under-estimates in two examples discussed in chapter 9 of *Thinking Clearly with Data*.

(a) In the discussion of violin practice, we noted that a musician with greater talent might both practice more and play the instrument better for reasons having nothing to do with how much she practices. Does this suggest that the correlation between practice and playing quality is an over-estimate or an under-estimate of the true effect of practice on playing?
(b) In the discussion of campaign spending, we argued that incumbents are likely to spend "heavily on their campaigns when they are electorally weak." Does this suggest that the observed lack of (or even negative) correlation between campaign spending and electoral performance of incumbents is an over-estimate or an under-estimate of the true effect of spending on votes?

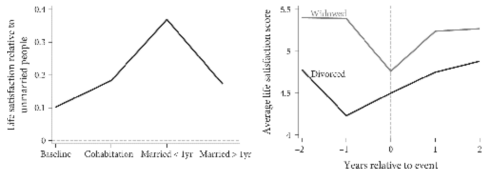
Time to put on a thinking cap for bonus points.

The question(s) below are all bonus. They are more time-consuming and difficult, so I'd recommend to only do them once you're satisfied with all previous answers.

Shortly after Harvard psychologist Daniel Gilbert's book, *Stumbling on Happiness*, was released, he was on TV, where he informed Stephen Colbert that "marriage is one of the best investments you can make in happiness." That advice implicitly rests on a causal claim: marriage causes happiness. Much recent research documents a positive correlation between marriage and happiness. But is the relationship causal?

(a) Provide an argument for why the correlation between marriage and happiness might be the result of reverse causation (happiness causing marriage, rather than the other way around).
(b) Identify two confounders that you think might make a causal interpretation of the correlation between marriage and happiness problematic. For each, explain why you believe the confounder might affect both treatment (being married) and outcome (happiness).
(c) Sign the bias for each of the confounders you identified. Having done so, explain whether each tends to make the observed correlation between marriage and happiness an over- or under-estimate of the true causal effect.
(d) A study by Anke Zimmermann and Richard Easterlin follows people from up to four years prior to their first marriage through several years after getting married. The basic finding is illustrated in the left-hand panel of the figure on this page, which shows the life satisfaction of people who got married during the study period relative to those who never got married during the study period. As we go from left to right, we see how the life satisfaction of a person changes over time as they first cohabitate with a partner, then get married, and continue that marriage for more than a year. Compare the life

satisfaction of people who have been married for a while to that of people who are not married but are living with their partner. Do you find this evidence supportive of or contrary to Gilbert's advice? Identify a confounder that this comparison suggests may have existed in the original correlation.



(e) A study by Jonathan Gardner and Andrew Oswald also follows individuals over time but asks a different question. It considers what happens to people's happiness when marriages end. The study looks at two ways a marriage might end: divorce or death of a spouse. The results are summarized in the right-hand panel of the figure. The horizontal axis shows years relative to an important event (divorce or widowhood) at time 0. The vertical axis shows life satisfaction. Life satisfaction is shown in black for those who became divorced and in gray for those who became widowed.

Notice the initial difference in life satisfaction between those who became widowed and those who got divorced, even before the event occurred. Does this difference make you more or less confident in Gilbert's causal interpretation? Why? Now consider the widows and widowers (gray line). How does their happiness change before, during, and after the year in which their spouses passed away? Does this make you more or less confident in Gilbert's causal interpretation? What does this comparison make you think might be going on in Gilbert's original correlation?