DATA1001 Topic 1 Flashcards

Design of Experiments

Abyan Majid

Question	My answer
What is data science?	Data science is a field that deals with exploration of data in order to unlock insights in particular domains.
What is a data scientist?	A data scientist is someone who can unlock insights from data that is applicable in the domain in which they work
What is domain knowledge?	Domain knowledge is a background understanding of the field in which the data scientist is operating.
What is a controlled experiment?	A controlled experiment is a study (to gather data) where the data scientist performs two parallel experiments, one of which is where the subjects are administered the treatment, while the other is where the subjects are administered a placebo (pretend treatment)
What are treatment and control groups?	Treatment group refers to subjects that receive the treatment, meanwhile control group refers to subjects that receive the placebo (pretend treatment)
What is confounding and what is bias? Differentiate!	Confounding refers to the influence of an extraneous, and very often hidden, on the relationship between the independent and dependent variables being studied. Meanwhile, bias refers to systematic errors in how data is gathered and used that affects the accuracy of the end result.
What is selection bias? Give one example!	Selection bias is bias wherein certain types of subjects are included in or excluded from the controlled experiment. It often occurs when either the subjects, investigators, or both, know the identities of the treatment and control groups.
What is a randomized controlled trial (RCT)? What problems can it solve and what can't it solve?	RCT is a strategy of assigning subjects into the treatment and control group purely by randomisation. It can solve selection bias to some extent in that the investigator has no control over who takes the treatment and who takes the placebo. However, it may not be purely random since it does not necessarily require that subjects or investigators be unaware of the identities of the treatment and control groups; observer bias may occur.
What is a placebo and what is a placebo effect? Give one example scenario involving a placebo and placebo effect!	A placebo is a pretend treatment ie. something else that is designed to mimic the treatment. The placebo effect is a phenomenon wherein a subject responds to the idea of the treatment (mostly because they

	thought they were taking the treatment). For example, in a study of whether or not coffee boosts attention span in lectures, someone who drank a placebo (but thinks they drank coffee) is more focused during this lecture than usual.
What is observer's bias? Give one example!	Observer's bias refers to the investigator intentionally discarding or reporting certain outcomes from the experiment. For example, in a study of whether a vaccine makes you immune to COVID-19, the investigator could commit observer's bias by choosing not to report outcomes wherein subjects who took the vaccine got diagnosed by COVID-19 afterwards.
What is a randomised controlled double-blind trial? Why is it considered gold-standard? How would we go about it and is it even possible?	A double-blind RCT is a randomized controlled trial wherein neither the subjects and investigators know about the identities of the treatment and control groups. It is called gold-standard because it would make for a purely random experiment with minimum room for biases and confounders. However, it is in most cases impossible. To go about attempting a double-blind RCT, we ought to have a 3rd party as an administrator of treatment and placebo in order to avoid observer's bias, and we ought to design the placebo to mimic the treatment as much as possible.
What is consent bias? Illustrate with an example of why avoiding consent bias might be ethically challenging!	Consent bias refers to subjects refusing to take the treatment/placebo, which implies that to avoid consent bias is to disregard consent in administering the treatment/placebo. For example, there is consent bias in letting subjects choose to take the drug or not. To avoid consent bias would mean to enforce rules as in who takes the drug and who does not. This is ethically questionable since it would mean that some sick people may not get the drug, and some healthy people may get the drug.
What is survivor bias? Give one example of survivor bias!	Survivor bias is a type of selection bias that occurs when certain subjects finish or do not finish the study. For example, if you use a dataset of billionaires, you'd conclude that the formula for success is to become a college dropout.
What is adherer bias? Give one example of adherer bias!	Adherer bias occurs when some subjects do not adhere to their supposed responsibilities in the experiment. For example, a sick patient who is supposed to consume a drug daily chooses not to consume it on some days.
What is an observational study?	An observational study is a study wherein the investigator has no control over the administration of treatment and placebo; assignment of subjects to the treatment and control groups is not done by randomization.

Can observational studies establish causation? Justify why or why not!	Observational can only establish association, not causation. This is due to the lack of randomisation and confounding variables (which there often appears to be many because you cannot control the study).
Why might an observational study be presented as a randomized controlled trial (RCT)?	An observational study may appear as a RCT because it may look as though the assignment of subjects into the treatment or control group is randomized when in fact it isn't. For example, the control group is historical but the treatment group is contemporaneous. In this case, time is a confounding variable.
What is a contemporaneous control group?	A control group that occurs at the same time as the treatment group
Give one example of a study and how you can control a confounder by splitting your subjects into subgroups!	In a study that explores the association between income and work hours per week, if we suspect that age is a confounding variable, we can split the subjects into multiple age groups to allow for potentially less biased comparison between income and work hours per week.
What is Simpson's paradox? And why can observational studies with a confounder lead to Simpson's paradox?	Simpson's paradox is a reversal of trend when subgroups of subjects are pooled together and it occurs as a result of not controlling for confounders. Confounding variables may have influence on the trends of the two variables being studied. For example, a study may report that income decreases with hours worked per week. However, when we split the subjects by a confounding variable: "field of occupation", we'd see the inverse ie. income increases with hours worked per week. Maybe we'd notice that construction workers work more hours than software engineers on average but are paid less.