Midterm Exam

Name and ID: _____

True or False questions.

1. In randomized experiments, association is causation.
2. Effect modifiers do not need to be independent with the treatment.
3. Sensitivity analysis is used to identify hidden confounders rather than to assess the impact of unmeasured confounding.
4. In a propensity score analysis, ignoring treatment effect heterogeneity across subgroups can lead to biased estimates.
5. The finer the stratification, the less the uncertainty introduced by random variability.
6. Controlling for measured confounders in a regression model always eliminates their biasing effects on causal inference.
7. Confounding is any systematic bias that would be eliminated by randomized assignment of treatment.
8. Selection bias can occur in both randomized controlled trials and observational studies, affecting internal validity.
9. For a faux marginal structural model $\mathbb{E}[Y^a|V] = \beta_0 + \beta_1 Va + \beta_3 V$ where V is a covariate, the parameter β_1 measures the causal effect of V on Y.
10. Sufficient cause interaction occurs when two component causes are both present in the same sufficient cause but do not interact with each other.

Multiple-Choice questions. (select all that apply.)

1. Which of the following are characteristics of a sufficient cause?
   (a) It consists of a complete set of component causes that leads to an outcome.
   (b) It remains sufficient even if any one of its component causes is removed.
   (c) Each component cause can also be present in other sufficient causes.
   (d) It requires all component causes to be present for the outcome to occur.
2. A potential outcome
   (a) is the same as a causal effect.
   (b) is not possible to estimate because very few individuals achieve their potential.
   (c) is the outcome for an individual under a potential treatment.
   (d) is the best outcome an individual can achieve.
3. Stratification requires
   (a) Positivity
   (b) exchangebility
   (c) Consistency
   (d) The Stable Unit Treatment Assumption (SUTVA)
4. Which methods can be used to adjust for time-varying confounders in longitudinal studies?
   (a) Marginal structural models
   (b) Propensity score weighting
   (c) G-estimation
   (d) Cox proportional hazards model
5. Which of the following scenarios can illustrate effect modification?
   (a) A treatment's effectiveness varies significantly between males and females.
   (b) An exposure's impact on an outcome differs between smokers and non-smokers.
   (c) A confounder influences the association between exposure and outcome by affecting both.
   (d) The treatment effect is modified by age, where it is stronger among younger individuals.

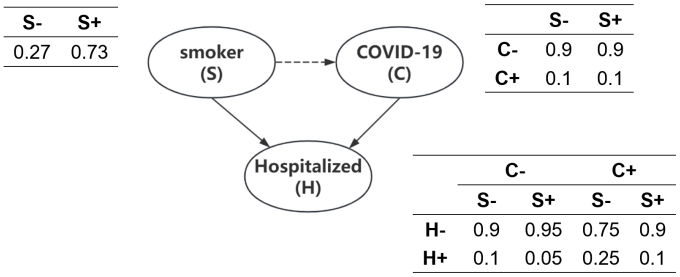
Short-Answer questions.

1. Please answer the following questions.

   (a) The following table summarizes the effect of a medical treatment on short-term recovery from a certain illness. Is the conclusion you infer from the whole population the same with those in stratified groups? If differ, explain why this paradox arises using DAG and elucidate how to avoid it in the analysis.

	All		Male		Female	
	Success	Failure	Success	Failure	Success	Failure
Treatment	20	20	8	5	12	15
Control	6	6	4	3	2	3

   (b) The following table gives a simple Bayesian Network structure. What would the conclusion be if we used only inpatient database? Is there any problem? If so, explain and elucidate how to avoid it in the analysis.



S-	S+
0.27	0.73

	S-	S+
C-	0.9	0.9
C+	0.1	0.1

	C-		C+	
	S-	S+	S-	S+
H-	0.9	0.95	0.75	0.9
H+	0.1	0.05	0.25	0.1

2. In a complete randomized experiment.

   (a) Describe the difference between Fisher's and Neyman's questions and concerns.

   (b) How does Fisher report the p-value? Prove its exactness.

   (c) How does Neyman report the p-value? Proving its unbiasedness in estimating the effect.

   (d) What if there is imbalance in baseline covariates? how will these two inferences change in the stratified randomized experiment?

3. The table below describes the result of a clinical trial. 50 participants were treated ($A = 1$) and 50 participants were untreated ($A = 0$). The primary outcome Y was cardiovascular death in 5 years. There are one covariate L (M for male and F for female).

# of participants	L	A	Y
20	F	0	0
5	F	0	1
10	F	1	0
10	F	1	1
15	M	0	0
10	M	0	1
10	M	1	0
20	M	1	1

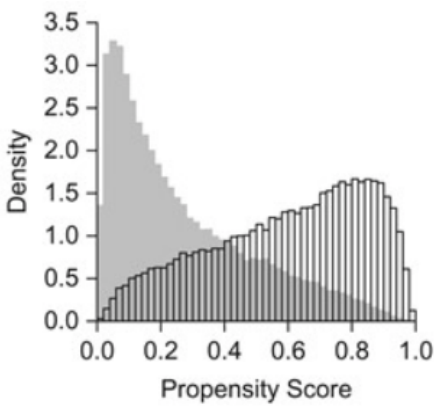
(a) Prove that the standardization and IP weighting are mathematically equivalent. Claim relevant assumptions made.

(b) In what circumstances would these two method be different? Claim the differences.

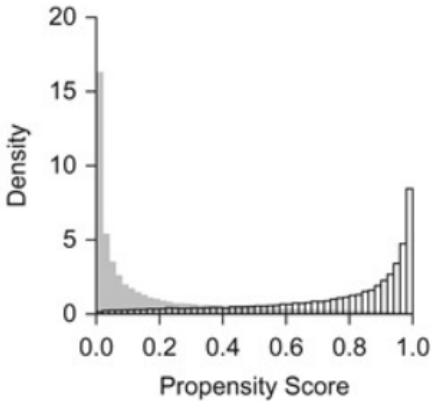
(c) Calculate the causal difference using stabilized IP weighting to adjust for the covariate L.

4. Now we are going to balance the potential confounders in a causal analysis.

   (a) we first used regression model to do propensity score analysis. We tried two different models. The following figures show the distribution of propensity scores. The unshaded bars indicate the control group; the gray shaded bars indicate the control group. What could we know from these figures and what model should we choose?



(a)



(b)

   (b) Here we have calculated propensity score $\hat{e}(x)$ for all participants (in the next page). How would you do matching to minimize the total distance between the control and case group (allowing multimatching)?

   (c) The outcome Y_i^{obs} is the systolic blood pressure for every participants after a month's medication. What is the estimate of the average treatment effect on the treated units (ATT) and control units (ATC)? What is the average treatment effects (ATE)?

Table 1: case group

i	$\hat{e}(x)$	Y_i^{obs}
1	0.46	116
2	0.20	110
3	0.34	120
4	0.57	125
5	0.19	105
6	0.48	118

Table 2: control group

i	$\hat{e}(x)$	Y_i^{obs}
1	0.32	120
2	0.28	115
3	0.42	123
4	0.47	135
5	0.54	140
6	0.52	130
7	0.20	112
8	0.19	135
9	0.15	120
10	0.22	140