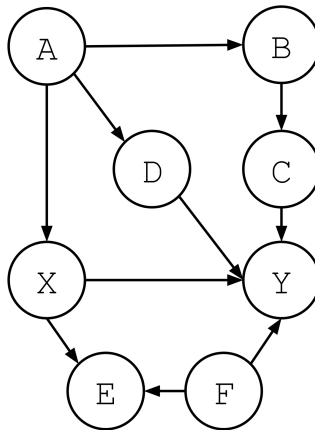# Ethics Part II - Week 3 Lab

Question Sheet

Zak Varty

Causal Networks, Colliders and Confounders

A principle investigator is designing a study to estimate the direct effect of a treatment X on outcome Y. They have constructed the following causal network, including measurable covariates A-F.

(a) Identify all causal paths between the treatment and outcome.

(b) By identifying colliders and confounders along these paths, explain which of these paths are open and which are closed when no covariates are controlled for.

(c) Identify the smallest set of variables to control for such that all backdoor paths are closed in this investigation.

(d) Identify a second, complementary, set of control variables that would also close all backdoor paths.

ATE and IPTW Estimation

MachineLearners is a start-up company developing educational software. Rather than teaching students by asking them questions in a predetermined, increasing level of difficulty, their software uses a machine learning algorithm to pose questions in a personalised order that is thought be of the greatest educational value to the student.

The MachineLearners team has conducted a trial to demonstrate the effectiveness of their system. They recruited students to their trial and asked them to take a baseline assessment. Each student in the trial was then assigned to either the control group or the treatment group. The control group were asked practice questions in a predetermined order, while the ordering of questions for the treatment group students was decided by the MachineLearner software. Finally, a follow up test was given to all students.

The results of these tests are given in the file `trial-data.csv`.

(a) Calculate a point estimate for the average treatment effect of the machine learner approach: that is, the expected improvement for a random individual if they practice with the MachineLearners ordering rather than the standard ordering.

(b) Using a non-parametric bootstrap approach, construct an approximate 95% confidence interval for the average treatment effect of the MachineLearners software.

A colleague at MachineLearners is concerned that the students recruited for the study might not be representative of their target market for the product. They realised that female students may be over represented in the study because of two popular single-sex schools in the area, both of which take only female students.

(c) Additional student information is contained in `school-data.csv`. Use this information and a non-parametric bootstrap approach to construct an approximate 95% confidence interval for the difference in the average treatment effect for male and female students. What implications does this have for the effectiveness of the MachineLearners system when used by their target demographic?

(d) The file `survey.csv` contains the results of a recent survey of the target market. Use the results of this survey and an appropriate weighting scheme to find a point estimate for the average treatment effect of the MachineLearners system, when used by students from their target market.

(e) Using a non-parametric bootstrap approach, construct an approximate 95% confidence interval for the treatment effect of the MachineLearners system, when used by students from their target market.