Name: ____ STAT 320: Midterm

This exam will focus on the causal effect of various weight loss strategies.

1. The causal effect of various weight loss strategies have been studied extensively, both with randomized experiments and observational studies.

   (a) (3 points) Give one reason a randomized experiment would be preferred to an observational study.

   Solution: Randomization balances all covariates (on average), observed and unobserved.

   (b) (3 points) Give one reason an observational study might be preferred to a randomized experiment in this context.

   Solution: Compliance might be an issue for studying weight loss strategies over extended periods of time; people may be more likely to stick to a strategy they choose themselves rather than one they were randomized to.

   (c) (8 points) When analyzing data from an observational study, give **four** ways we can make the observational study more similar to a randomized experiment.

   Solution: There are many possible answers, some possible answers listed here:

   1. Remove outcomes for the entire design phase (balancing covariates)
   2. Do as much as possible in the *design* phase, so the analysis is easy and straightforward
   3. Try to balance covariates between treatment groups, for example using propensity score subclassification, matching, or weighting
   4. In a randomized experiment the assignment mechanism is known; in an observational study we can try to model it by estimating the propensity score
   5. A randomized experiment is probabilistic by design; try to make an observational study probabilistic by eliminating units with propensity scores close to 0 or 1
   6. A randomized experiment is unconfounded by design; try to make an observational study unconfounded by collecting data on and balancing important covariates

2. (2 points) Our outcome is weight change over the course of the year. Why might we want to use *weight change* as an outcome, rather than just weight at the end of the year?

> **Solution:** Less variability in weight change than in final weight; improve precision of causal effect estimates

3. (3 points) All data in this exam is from the 2007-2008 NHANES (the National Health and Interview Examination Survey). We restrict our analysis to people aged 16 and older who answered yes to the question "During the past 12 months, have you tried to lose weight?," giving $n = 1746$. Why do we restrict our analysis to those who answered yes to this question?

> **Solution:** To help make it probabilistic. If somone was not trying to lose weight, they wouldn't have had any chance of using a particular weight loss strategy.

4. *Weight Loss Program - Observational Study*
For this question, the treatment variable is whether a person joined a formal weight loss program (such as Weight Watchers, Jenny Craig, Tops, or Overeaters Anonymous) in the 12 months prior to the survey. You may assume that all of these weight loss programs focus exclusively on diet. About 5% of the NHANES sample had joined a weight loss program in the past year.

   (a) (8 points) Below is a subset of some of the other variables included in the survey. The first four are at the time of the survey, the latter four were answered at the time of the survey but were asked about the past 12 months. Please check the ones that you believe should be included as covariates.

   $\surd$ **Age**
   $\surd$ **Gender**
   ◯ BMI
   $\surd$ **Height**
   ◯ Average minutes of exercise per week during the past year
   ◯ Whether they tried to eat less calories in the past year
   ◯ Whether they tried to eat less fat in the past year
   ◯ Whether they tried to eat less carbs in the past year

(b) (4 points) Of the three techniques we've learned for balancing covariates, which of the following do you think would be best in this case? Briefly explain your answer.
○ Subclassification √ **Matching** ○ Weighting

Solution: The control group is much larger than the treatment group (only 5% of units are treated), making this an ideal situation for matching.

(c) (3 points) Based on your answer to b, define your *estimand*, using either words or notation.

Solution: ATT: Average treatment effect for the treated

(d) (10 points) Regardless of your answer to b, the first step is to estimate propensity scores. Describe the process you would use to estimate propensity scores. There are a set of primary covariates which are known to be important regarding weight loss, and a set of secondary covariates which may or may not be important.

Solution: Use logistic regression with treatment assignment as the response, and at least all primary covariates as predictors. Use some type of variable selection (such as stepwise regression or likelihood ratio tests) to decide whether to include secondary covariates, second-order interactions of primary covariates, and potential transformations of continuous primary covariates.

Once a model has been decided upon, estimate propensity scores, and trim non-overlapping units (eliminate controls with estimated propensity scores below the lowest treated, and treated with propensity scores above the highest control). Refit model and iterate until overlapping. Use this final model to estimate propensity scores for each unit.

5. *Weight Loss Program - Randomized Experiment*
Suppose a completely randomized experiment was conducted where the goal was to answer the question "Does joining a weight loss program cause weight loss?"

(a) (5 points) Describe how Fisher would answer this question.

Solution: Supposing the sharp null hypothesis of no treatment effect is true ($H_0 : Y_i(0) = Y_i(1)$ for all i), generate a randomization distribution by rerandomizing units to treatment groups and calculating the test statistic assuming the sharp null (or by calculating the test statistic under the null for all possible randomizations), and then calculate a p-value as the proportion of these simulated statistics that are as extreme as your observed statistic.

(b) (5 points) Describe how Neyman would anwer this question.

Solution: Calculate the estimate (observed difference in means) and the estimated variance of this estimate:

$$\hat{var}(\hat{\tau}) = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2},$$

and then assuming sample sizes are large enough, use the t-distribution to calculate a p-value against the null hypothesis of no average treatment effect, and a confidence interval for the true difference in means.

(c) (4 points) You have data on initial obesity (obese or not), and believe this covariate to be very important. If you could design the experiment, what's one way you could ensure this covariate is balanced when you *design* the experiment? (be specific about what you would do - don't just give a name of a technique)

Solution: Do a stratified randomized experiment; randomize people to participate in a weight loss program or not separately within obese people and non-obese people.

(d) (4 points) If the experiment has already been conducted, and initial obesity happens to be imbalanced, what's one way you could account for this imbalance *in analysis*? (be specific about what you would do - don't just give a name of a technique)

Solution: Include the covariate obese or not, X, in the regression model $Y_i^{obs} = \alpha + \tau W_i + \beta X_i + \varepsilon_i$.

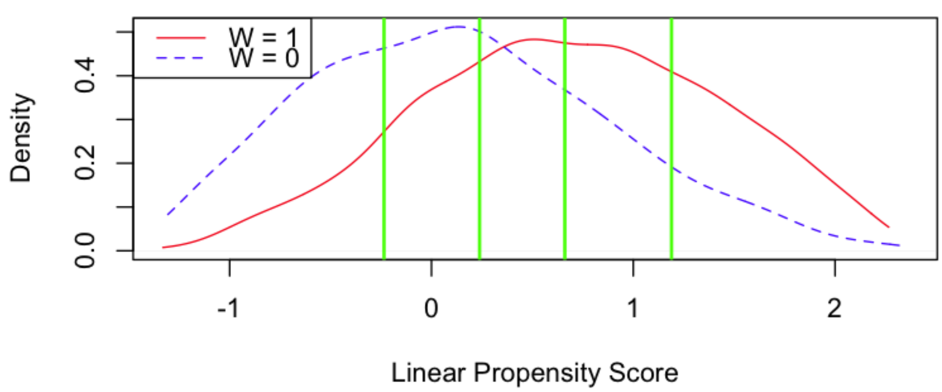

Figure 1: Linear propensity scores by treatment group with subclass breaks.

6. For this question, we return to the NHANES data, but this time the treatment variable is whether a person had exercised regularly in the past year (whether their average minutes of exercise per week for the past year was greater than 0). For simplicitly, we'll call these people "exercisers," as compared to non-exercisers. About 60% of the sample were exercisers. We estimate the propensity scores for a trimmed sample, and then subclassify based on the linear propensity score, using 5 equally sized subclasses.

   (a) (2 points) What is the "linear propensity score"?

   Solution: The log-odds of the propensity score:

   $$\log\left(\frac{e(x)}{1-e(x)}\right)$$

   (b) (2 points) The linear propensity scores are shown in Figure 5. Why does the solid curve have linear propensity scores higher than the dashed curve?

   Solution: The solid curve represents the treated group, and the propensity scores are probability of being treated, so should be higher in the treated group.

   (c) (3 points) Below are the sample sizes in the subclasses:

```
subclass
    1   2   3   4   5
0 232 167 129 103  60
1 114 178 216 242 285
```

   Based on this information, Figure 1 and Figure 2 (next page), do you think more subclasses should be used? Why or why not?

   Solution: Based on the love plot, covariate balance looks very good with these subclasses, so further subclassification would only decrease sample size in each subclasses, and so decrease precision.

   (d) (2 points) Based on Figure 2, explain how the closed (solid) dot for age was calculated (or what the dot represents). Be more specific than just "balance".

   Solution: The t-statistic for age; observed difference in mean age between treatment and control groups divided by the standard error

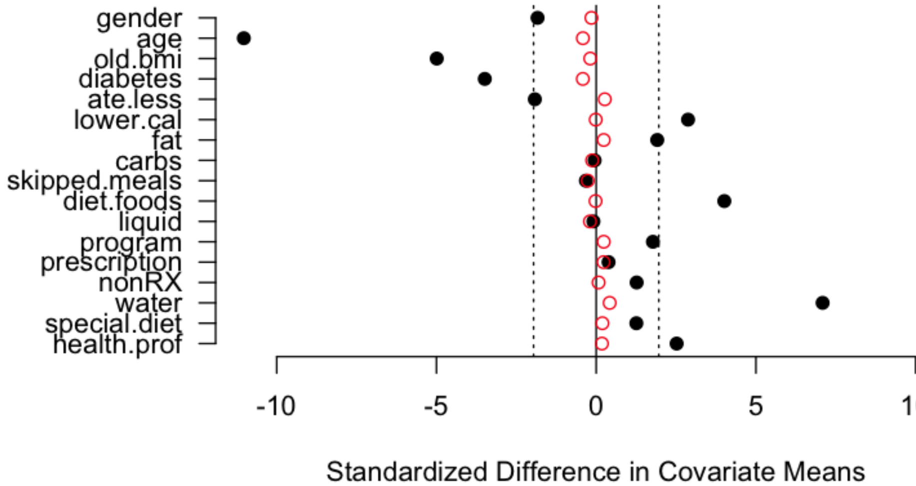

Figure 2: Love plot: solid dots represent balance (treatment - control) before subclassification, open dots represent balance after subclassification.

(e) (3 points) Interpret the initial imbalance for age in the context of the study.

> **Solution:** People in the treatment group (exercisers) were more than 10 standard errors younger than people in the control group (non-exercisers), on average.

(f) (4 points) Explain how the open dot for age in Figure 2 was calculated. Address the numerator **and** the denominator.

> **Solution:** The numerator is the weighted average difference in mean age after subclassification; an average of the difference in mean age within each subclass, weighted by size of the subclass. The denominator is the same denominator as that used for the black solid dots, the standard error of the original difference in mean age.

(g) (3 points) All covariates included in the propensity score model are shown on the love plot. Do you think the study is unconfounded? Why or why not?

> **Solution:** No; the study does not include all relevant covariates. For example, how physically fit a person was at the beginning of the study probably differs between the people that exercised in the past year and those who didn't, and probably also effects weight change.

7. *Analysis*

(a) (2 points) The average difference in observed mean weight change (exercisers - non-exercisers) within each subclass is given below. Estimate the average causal effect of exercise on weight change.

```
-1.42 -4.28  1.82 -1.20 -3.57
```

Solution: These subclasses are equally sized, so we take a simple average of the estimates within each subclass: -1.73.

(b) (4 points) The standard errors for these estimates within each subclass are given below. Estimate the standard error of your estimate in a.

```
1.81 1.80 1.88 2.13 2.86
```

Solution: To find the variance we add up the standard error of each subclass squared, times the squared weight for that subclass $(1/5)^2$, and then take the square root to get the standard error:

$$\sqrt{\left(\frac{1}{5}\right)^2 (1.81^2 + 1.8^2 + 1.88^2 + 2.13^2 + 2.86^2)} = 0.95$$

(c) (4 points) Based on your answers to a and b, give and interpret a 95% confidence interval for the average causal effect of exercise on weight change.

Solution:

$$-1.73 \pm 2 \times 0.95 = (-3.63, 0.17)$$

We are 95% confident that joining a weight loss program people's weight to change somewhere between losing 3.63 pounds and gaining 0.17 pounds over the course of a year, on average.

(d) (2 points) Based on your answers to a and b, does this analysis provide evidence that exercising causes weight loss? Why or why not?

Solution: No; the confidence interval includes 0. Or you could compute the t-statistic as $t = -1.73/0.95 = -1.82$, and because this is less than 2 in absolute value we know it is not significant at $\alpha = 0.05$.

(e) (3 points) Suppose we wanted to further balance covariates within each subclass, so estimate the treatment effect within each subclass using regression. Give the model you would use within each subclass.

Solution: $Y_i^{obs} = \alpha + \tau W_i + \beta' X_i + \varepsilon_i$

(f) (2 points) Why is it better (safer) to do regression within subclasses rather than regression on the entire data to estimate the causal effect?

Solution: Covariates are better balanced, so model is not so reliant on extrapolation

8. (2 points) What do you think is the most valuable thing you have learned so far in this course, and why?

Solution: Answers will vary.