Antonella Basso

PHP 2610 Problem Set 1

Due: September 23 by 11:59pm

Instructions: Please upload your answer to the Canvas course page as a pdf file. You can submit your answers in a separate pdf file (please be sure to properly mark the question number to your responses), or you can work on this pdf file, scan it, and upload it to the Canvas course page.

Late or missed assignments: Problem sets and the final report must be turned in online at or before the posted due date. Every one day (24 hours) of delay will result in a ten point (out of 100) downgrade.

Question 1 (30 points)

Please read the abstract from the following paper and then answer questions about it. It is not necessary to read the full paper to answer the questions below.

Harter, P., Sehouli, J., Vergote, I., Ferron, G., Reuss, A., Meier, W., ... & du Bois, A. (2021). Randomized trial of cytoreductive surgery for relapsed ovarian cancer. *New England Journal of Medicine, 385*(23), 2123-2131.

Background: Treatment for patients with recurrent ovarian cancer has been mainly based on systemic therapy. The role of secondary cytoreductive surgery is unclear.

Methods: We randomly assigned patients with recurrent ovarian cancer who had a first relapse after a platinum-free interval (an interval during which no platinum-based chemotherapy was used) of 6 months or more to undergo secondary cytoreductive surgery and then receive platinum-based chemotherapy or to receive platinum-based chemotherapy alone. Patients were eligible if they presented with a positive Arbeitsgemeinschaft Gynäkologische Onkologie (AGO) score, defined as an Eastern Cooperative Oncology Group performance-status score of 0 (on a 5-point scale, with higher scores indicating greater disability), ascites of less than 500 ml, and complete resection at initial surgery. A positive AGO score is used to identify patients in whom a complete resection might be achieved. The primary end point was overall survival. We also assessed quality of life and prognostic factors for survival.

Results: A total of 407 patients underwent randomization: 206 were assigned to cytoreductive surgery and chemotherapy, and 201 to chemotherapy alone. A complete resection was achieved in 75.5% of the patients in the surgery group who underwent the procedure. The median overall survival was 53.7 months in the surgery group and 46.0 months in the no-surgery group (hazard ratio for death, 0.75; 95% confidence interval, 0.59 to 0.96; p-value=0.02). Patients with a complete resection had the most favorable outcome, with a median overall survival of 61.9 months. A benefit from surgery was seen in all analyses in subgroups according to prognostic factors. Quality-of-life measures through 1 year of follow-up did not differ between the two groups, and we observed no perioperative mortality within 30 days after surgery.

Conclusions: In women with recurrent ovarian cancer, cytoreductive surgery followed by chemotherapy resulted in longer overall survival than chemotherapy alone.

Q1.1 (7 points) What are the units in this example?

a. Surgeons who underwent cytoreductive surgery
b. Patients who were assigned to cytoreductive surgery and chemotherapy

c. Patients who were assigned to chemotherapy only
d. Patients without ovarian cancer at baseline
e. Patients with recurrent ovarian cancer who are eligible for randomization

Answer: E

Q1.2. (7 points) What are the treatment and control conditions in this example?

a. Treatment: Cytoreductive surgery and chemotherapy, Control: chemotherapy only
b. Treatment: Cytoreductive surgery and no chemotherapy, Control: chemotherapy only
c. Treatment: Cytoreductive surgery and chemotherapy, Control: Cytoreductive only
d. Treatment: Cytoreductive surgery, Control: no related therapy
e. Treatment: Chemotherapy, Control: no related therapy

Answer: A

Q1.3. (8 points) What are the two potential outcomes using the primary end point in this example?

a. $Y^{A=1}$: survival rate if patients chose to receive cytoreductive surgery and chemotherapy, $Y^{A=0}$: survival rate if patients chose to chemotherapy alone
b. $Y^{A=1}$: survival rate if patients chose to receive cytoreductive surgery alone, $Y^{A=0}$: survival rate if patients chose to chemotherapy alone
c. $Y^{A=1}$: survival rate if patients were assigned to cytoreductive surgery and chemotherapy, $Y^{A=0}$: survival rate if patients were assigned to chemotherapy alone
d. $Y^{A=1}$: survival rate if patients were assigned to cytoreductive surgery alone, $Y^{A=0}$: survival rate if patients were assigned to chemotherapy alone
e. $Y^{A=1}$: survival rate if patients were assigned to cytoreductive surgery and chemotherapy, $Y^{A=0}$: survival rate if patients were choose to chemotherapy alone

Answer: C

Q1.4. (8 points) What is the correct interpretation of the average treatment effect (ATE)?

a. The difference in average outcomes between a group that receives the treatment and a group that receives the control
b. The difference in average outcomes if a whole population is given the treatment compared to their average outcomes if that same group is given the control
c. The difference in average outcomes if a whole population is given the treatment compared to their average outcomes if that same group is assigned to either the control or the treatment
d. The difference in average outcomes if the treated population is given the treatment compared to their average outcomes if that same group is given the control
e. The difference in average outcomes between a group that is randomly assigned to the treatment or the control and a group that receives the control.

Answer: B

Question 2 (30 points)

This assignment uses observational data to examine the effect of a labor training program on real earnings in 1978. The data (`rhc.csv`) is available at the Canvas course page (Files/Data). The variables (1 outcome, 1 treatment, 10 covariates) in our data set are:

- Units: ICU patients in 5 hospitals

- Outcome : binary indicator for death (yes/no) (`'died'`)
- Treatment: right heart catheterization (rhc) vs. not (`'treatment'`)

Q2.1. (4 points) Let us first consider two ways to estimate the treatment effect and the associated standard 95% confidence intervals using treated and control subjects. The first method is the Fisher's Exact estimation of the odds ratio of deaths in the treatment group relative to the odds of deaths in the control group. Please calculate the odds ratio using `fisher.test` function (round to 2 decimal places):

Answer: **1.25**

Q2.2. (4 points) Please derive the 95% confidence interval associated with the Fisher's Exact Test estimation of the above odds ratio (round to 2 decimal places and answer in the form (L, U), where L and U indicate the lower and upper bounds, respectively):

Answer: **(1.12, 1.40)**

Q2.3. (8 points) The second method is a regression-adjusted estimation using a logistic regression of the outcome on the treatment indicator and predictors. Please provide the same odds ratio using the regression adjustment that conditions on 10 covariates (no interaction effects included in the model).

Answer: **1.37**

Q2.4. (8 points) Please derive the 95% confidence interval associated with the regression adjustment of the above odds ratio (round to 2 decimal places and answer in the form (L, U), where L and U indicate the lower and upper bounds, respectively):

Answer: **(1.21, 1.55)**

Q2.5. (6 points) Which covariates do you think are particularly important to obtain good balance on when estimating the treatment effect using this data (Choose 3)?

(a.) MOSF
b. Colcan
(c.) COMA
(d.) Female
e. Age

Answer: **A, C, D**

Question 3 (40 points)

Please use the same data set from Question 2. Now consider propensity score weighting methods to estimate the treatment effect using `WeightIt` and `MatchIt` packages. Include all 10 covariates (as a main effect; no interaction terms included) in the propensity score model.

Q3.1. (4 points) What are the minimum (min, max) and maximum (min, max) of the estimated propensity score weights when the target estimand is ATT?

(a.) Treatment group (min, max) = (1.00, 1.00), Control group (min, max) = (0.13, 1.52)
b. Treatment group (min, max) = (1.00, 1.52), Control group (min, max) = (0.13, 1.52)
c. Treatment group (min, max) = (1.66, 8.88), Control group (min, max) = (1.00, 1.00)
d. Treatment group (min, max) = (1.66, 8.88), Control group (min, max) = (1.13, 2.52)

e. Treatment group (min, max) = (1.00, 1.00), Control group (min, max) = (1.00, 1.00)

Answer: A

Q3.2. (6 points) Please derive the odd ratio of deaths in the treatment group relative to the odds of deaths in the control group using the ATE-weighted outcome regression. Use `survey` package with `family = quasibinomial()` (round to 2 decimal places).

Answer: 1.40

Q3.3. (6 points) Please derive the 95% confidence interval for the above odds ratio (round to 2 decimal places and answer in the form (L, U), where L and U indicate the lower and upper bounds, respectively).

Answer: (1.24, 1.59)

Q3.4. (6 points) Please derive the odds ratio of deaths in the treatment group relative to the odds of deaths in the control group using the ATT-weighted outcome regression. Use `survey` package with `family = quasibinomial()` (round to 2 decimal places).

Answer: 1.35

Q3.5. (6 points) Please derive the 95% confidence interval for the above estimate (round to 2 decimal places and answer in the form (L, U), where L and U indicate the lower and upper bounds, respectively).

Answer: (1.19, 1.53)

Q3.6. (6 points) Please derive the odds ratio of deaths in the treatment group relative to the odds of deaths in the control group using 1:1 nearest neighbor matching. Use `family = quasibinomial()` in the outcome model (round to 2 decimal places).

Answer: 1.36

Q3.7. (6 points) Please derive the 95% confidence interval for the above estimate (round to 2 decimal places and answer in the form (L, U), where L and U indicate the lower and upper bounds, respectively).

Answer: (1.19, 1.55)

Code

```
# libraries
library(tidyverse)
library(tableone)

library(cobalt)

library(Matching)
library(MatchIt)

library(WeightIt)
library(survey)

library(readxl)
library(multcomp)

# importing "amenorrhea" data
rhc <- read.csv("/Users/antonellabasso/Desktop/PHP2610/Data/rhc.csv")
head(rhc)
```

```
####### Q2.1 - Q2.2 #######

# frequency table
## cols: treatment, no treatment
## rows: death, no death
freq_death_treat <- matrix(c(nrow(rhc[rhc$treatment==1 & rhc$died==1,]), # died w/ rhc
                             nrow(rhc[rhc$treatment==0 & rhc$died==1,]), # died w/o rhc
                             nrow(rhc[rhc$treatment==1 & rhc$died==0,]), # lived w/ rhc
                             nrow(rhc[rhc$treatment==0 & rhc$died==0,])), # lived w/o rhc
                           nrow=2, ncol=2, byrow=TRUE)

## odds ratio: (1486*1315)/(698*2236) = 1.25

# Fisher's Exact Estimation
## odds ratio of deaths under treatment relative to no treatment (control)
fisher.test(freq_death_treat) # OR: 1.25; 95% CI: (1.12, 1.40)
```

```
####### Q2.3 - Q2.4 #######

# logistic regression (of outcome on all predictors)
RAE_OR <- glm(died ~ ., family=binomial(link="logit"), data=rhc)

# regression-adjusted estimation of odds ratio
exp(coef(RAE_OR))["treatment"] # OR: 1.37
exp(confint(RAE_OR)) # 95% CI: (1.21, 1.55)


####### Q2.5 #######

# statistically signif. predictors of outcome:
## Cirr, Coma, lungcan, MOSF, age, treatment (p-value < 0.001)
## sepsis, female (p-value < 0.05)
summary(RAE_OR)
```

```r
# statistically signif. predictors of treatment:
## ARF, CHF, Coma, MOSF, sepsis, female (p-value < 0.001)
## Cirr (p-value < 0.01)
summary(glm(treatment ~ . -died, family=binomial(link="logit"), data=rhc))

## chosen for balance: MOSF, Coma, female
```

```r
####### Q3.1 #######

# propensity score weighting to estimate treatment effect

## propensity score model - ATT
ps_model_ATT <- weightit(treatment ~ . -died, data=rhc,
                         estimand="ATT", method="ps")

## (control) ps weight summaries
summary(ps_model_ATT) # (min, max) = (0.13, 1.52)
```

```r
####### Q3.2 - Q3.3 #######

# propensity score model - ATE
ps_model_ATE <- weightit(treatment ~ . -died, data=rhc,
                         estimand="ATE", method="ps")

# ATE-weighted outcome regression
design_w <- svydesign(~ 1, weights=ps_model_ATE$weights, data=rhc)
fit_w <- svyglm(died ~ ARF + CHF + Cirr + colcan + Coma + lungcan +
                  MOSF + sepsis + age + female + treatment,
                family=quasibinomial(), design=design_w, data=rhc)
summary(fit_w)

## estimates
exp(coef(fit_w))["treatment"] # OR: 1.40
exp(confint(fit_w)) # 95% CI: (1.24, 1.59)
```

```r
####### Q3.4 - Q3.5 #######

# ATT-weighted outcome regression
design_w2 <- svydesign(~ 1, weights=ps_model_ATT$weights, data=rhc)
fit_w2 <- svyglm(died ~ ARF + CHF + Cirr + colcan + Coma + lungcan +
                   MOSF + sepsis + age + female + treatment,
                 family=quasibinomial(), design=design_w2, data=rhc)
summary(fit_w2)

## estimates
exp(coef(fit_w2))["treatment"] # OR: 1.35
exp(confint(fit_w2)) # 95% CI: (1.19, 1.53)
```

```r
####### Q3.6 - Q3.7 #######

# 1:1 nearest neighbor matching
## model
nnm_model <- matchit(treatment ~ ARF + CHF + Cirr + colcan + Coma + lungcan +
                   MOSF + sepsis + age + female, data=rhc,
```

```
                    estimand="ATT", method="nearest")

summary(nnm_model, standardize=TRUE)

## fit
nnm_fit <- glm(died ~ ARF + CHF + Cirr + colcan + Coma + lungcan +
                 MOSF + sepsis + age + female + treatment,
               family=quasibinomial(), data=match.data(nnm_model))

## estimates
exp(coef(nnm_fit))["treatment"] # OR: 1.36
exp(confint(nnm_fit)) # 95% CI: (1.19, 1.55)
```