EDS241: FINAL

Hope Hahn

02/23/2024

Make sure to read through the setup in markdown. Remember to write out interpretations and report your results in writing/table/plot forms.

1 Part 1: RCTs, treatment ignorability (selection on observables), propensity scores *(15 points total)*

Setup

This exercise is inspired by Costello et al. 2008 article in science "Can Catch Shares Prevent Fisheries Collapse", which we also discussed in class (lecture 5). "Inspired" means that the data final_fisheries_data.csv are synthetically generated to simplify things for our purposes. It contains the variables on 11,135 fisheries (only cross sectional, no time observations): These fisheries were either regulated by an Individual Transferable Quota (ITQ) for all years between 1990 and 2012 or in none of those years. Variables in the dataset include:

The outcome and treatment variables are:

COLL_SHARE = share of years a fishery is collapsed between 1990 and 2012 (collapse defined as harvest being more than 10% below maximum recorded harvest).

ITQ = dummy variable indicating 'treatment' with an ITQ (equal to 1 if the fishery has been regulated by an ITQ and 0 otherwise).

The control variables are:

MET1, MET2, ...MET6 = Dummy variables indicating to which Marine Ecosystem Type (MET) the fishery belongs to (coral reefs, kelp forests, seagrass meadows, open ocean, deep sea, mangrove forests). This type does not change over the relevant time period and does not depend on human influence.

IND_SR = Index of species richness in 1980 with values between 0 and 100 indicating the biodiversity with respect to species in the fishery. Bounds of 0 and 100 are the lowest and highest observed values of species diversity across all fisheries in 1980, respectively.

COMM_VAL = Commercial value of fisheries in 1980 in million US-$

The basic question of interest is "What is the average treatment effect of implementing an ITQ in the time period from 1990 to 2012 on the share of years with a collapse. It is likely that the probability a fishery is selected for an ITQ depends on the pre-treatment characteristics given. It is also quite likely that the pre-treatment characteristics have an effect on the share of collapse for each fishery, i.e. our outcome variable of interest.

Question ().1 Setup Code

To prepare the data for the rest of the analysis, I combined all MET categories into a single column. This was mostly to be able to plot back to back in one plot.

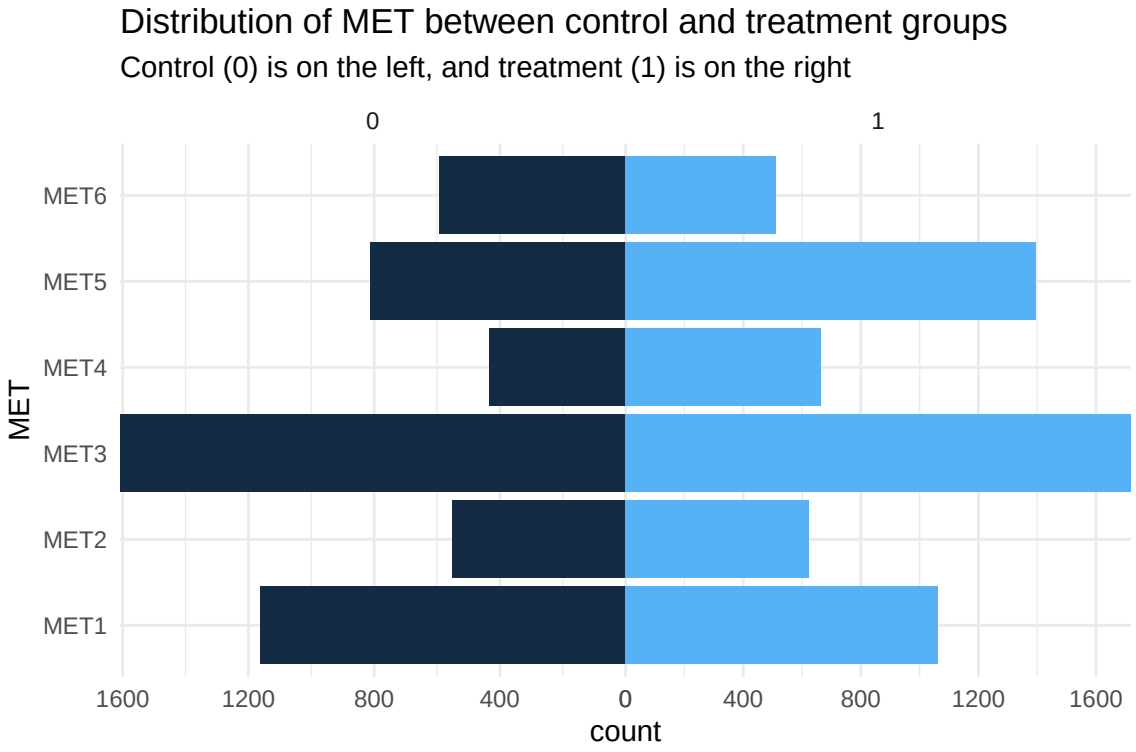
```
rm(list=ls()) # clean environment

## Load Data
fisheries_data <- read_csv(here::here("final", "data", "final_fisheries_data.csv"))

# this data has combined MET column
fisheries_data_new <- fisheries_data %>%
  mutate(MET1 = ifelse(MET1 == 1, "MET1", NA), # these rows turn values of 1 into the category of MET
         MET2 = ifelse(MET2 == 1, "MET2", NA),
         MET3 = ifelse(MET3 == 1, "MET3", NA),
         MET4 = ifelse(MET4 == 1, "MET4", NA),
         MET5 = ifelse(MET5 == 1, "MET5", NA),
         MET6 = ifelse(MET6 == 1, "MET6", NA)
  ) %>%
  mutate(MET = coalesce(MET1, MET2, MET3, MET4, MET5, MET6)) %>% # combine all METS into a single varia
  select(MET, IND_SR, COMM_VAL, ITQ, COLL_SHARE) # select for all relevent variables (remove all the ME
```

Question (a) Pretreatment Ecosystem Characteristic Comparison, Visual *(3 pts)*

(a) Compare the distributions of pre-treatment ecosystem characteristics (i.e. MET1, MET2, „, MET6) between the treated and the control groups by drawing back to back histograms [2 pts]. Write one sentence discussing the (dis)similarity between the two groups [1pt].

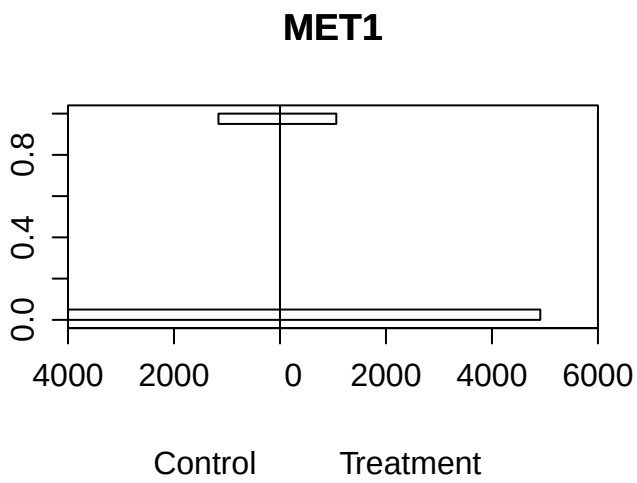
Back to back plot showing distribution of all marine ecosystem types accross treatment and control groups

```
# bar data
bar_data <- fisheries_data_new %>%
  mutate(count = 1)

# make control group negative
bar_data$count[bar_data$ITQ == '0'] <- -(bar_data$count[bar_data$ITQ == '0'])


# attempting to put all variables in one graph
ggplot(bar_data, aes(x = MET, y = count, fill = ITQ)) +
  geom_col(show.legend = FALSE) + # remove legend
  facet_wrap(~ ITQ, scales = "free_x") + # facet by treatment group
  coord_flip() + # flip coordinates
  scale_y_continuous(
    expand = c(0, 0), # edit y scale
    labels = function(x) signif(abs(x), 3)) + # make the counts all positive
  labs(title = "Distribution of MET between control and treatment groups", # update title/subtitle
       subtitle = "Control (0) is on the left, and treatment (1) is on the right") +
  theme_minimal() + # make theme lighter
  theme(panel.spacing.x = unit(0, "mm")) # make plots flush with each other
```

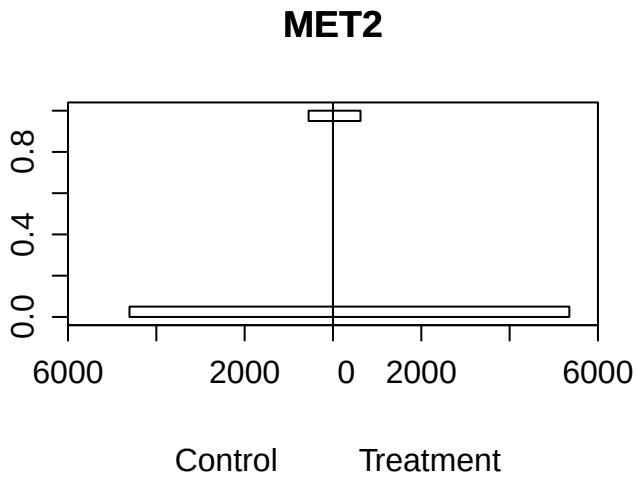


Back to back plots showing distribution of marine ecosystem type distribution on individual plots

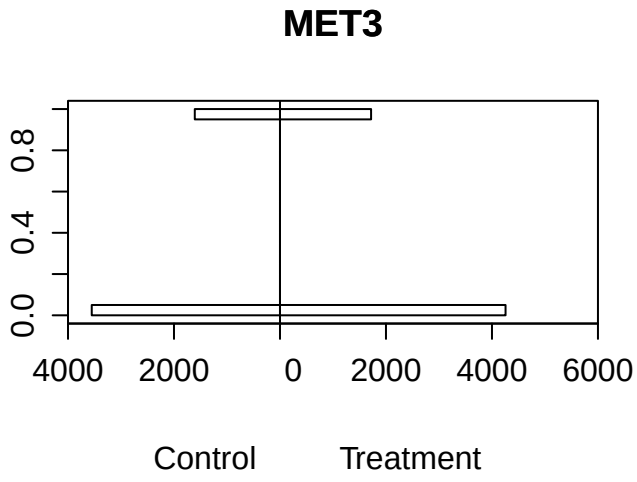
```
## Histograms comparing covariates met1
histbackback(split(fisheries_data$MET1, fisheries_data$ITQ),
    main = "MET1", xlab = c("Control", "Treatment"))
```



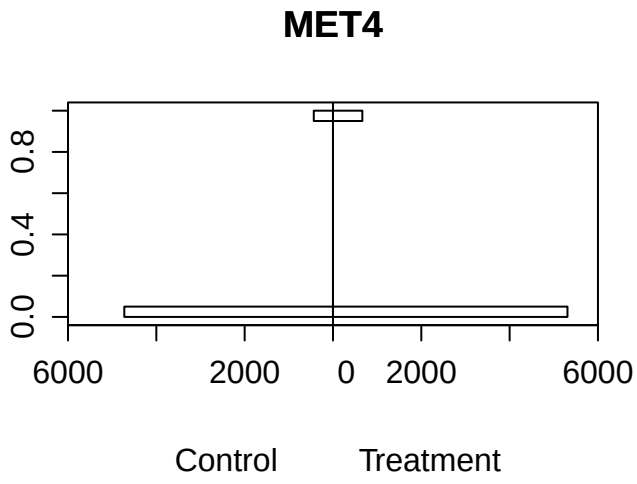

```
# met2
histbackback(split(fisheries_data$MET2, fisheries_data$ITQ),
    main = "MET2", xlab = c("Control", "Treatment"))
```



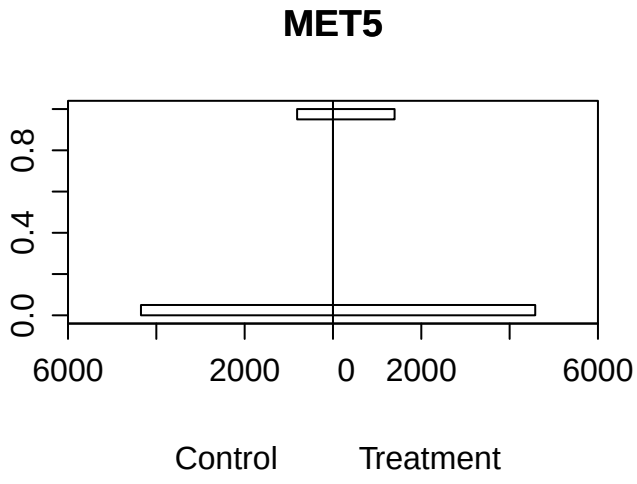

```
# met3
histbackback(split(fisheries_data$MET3, fisheries_data$ITQ),
    main = "MET3", xlab = c("Control", "Treatment"))
```



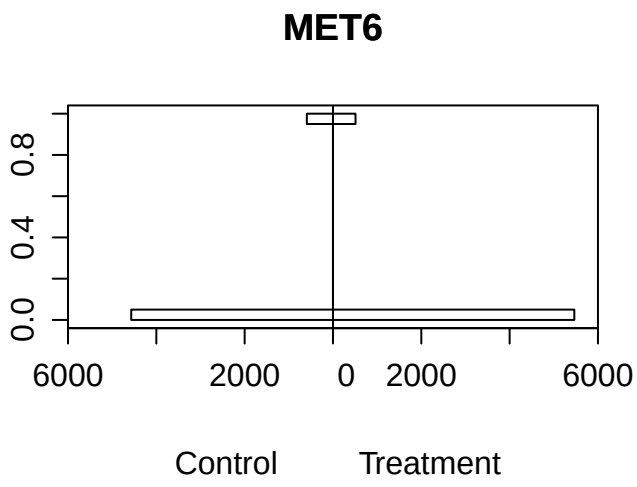

```
# met4
histbackback(split(fisheries_data$MET4, fisheries_data$ITQ),
    main = "MET4", xlab = c("Control", "Treatment"))
```




```
# met5
histbackback(split(fisheries_data$MET5, fisheries_data$ITQ),
    main = "MET5", xlab = c("Control", "Treatment"))
```




```
# met6
histbackback(split(fisheries_data$MET6, fisheries_data$ITQ),
    main = "MET6", xlab = c("Control", "Treatment"))
```



- **For each MET variable (1-6) the histograms show that each marine ecosystem type are relatively equally divided between the groups regulated by an ITQ (treatment) and unregulated (control). However, when looking at the first graph containing all MET, it appears that MET4-6 are less equally divided (have more imbalance) among treatment and control groups. Without significance tests, it is difficult to determine whether these differences are significant between the two groups, and we are not able to definitively know whether there is a significant covariate imbalance among the marine ecosystem types; we can only make assumptions visually.**

Question (b) Pretreatment Ecosystem Characteristic Comparison, Mean differences *3 pts)*

(b) Do a test on mean differences between the treated and control groups for the species richness index (IND_SR) and commercial value (COMM_VAL) variables. Interpret the results (estimated difference and significance) [2 pts] and make a conclusion regarding the similarity between the groups [1pt].

```
# split treat and control to diff datasets
fisheries_treat <- fisheries_data_new %>%
    filter(ITQ == 1)
fisheries_control <- fisheries_data_new %>%
    filter(ITQ == 0)

# find mean difference between two groups
difference_IND_SR <- round((mean(fisheries_treat$IND_SR) - mean(fisheries_control$IND_SR)),
    2)
difference_COMM_VAL <- round((mean(fisheries_treat$COMM_VAL) -
    mean(fisheries_control$COMM_VAL)), 2)

# add mean differences into little data frame
difference_df <- data.frame(Variable = c("IND_SR", "COMM_VAL"),
    mean_difference = c(difference_IND_SR, difference_COMM_VAL))

# t-tests of continuous variables select continuous
```

```r
# variables
continuous_vars <- fisheries_data %>%
    select(IND_SR, COMM_VAL, ITQ)

# initialize empty data frame
t_test_results <- data.frame()

# select the names of continuous variables
continuous_names <- names(continuous_vars)[1:2]

# for loop for t-test
for (var in continuous_names) {

    formula <- as.formula(paste(var, "~ ITQ"))

    # t-test
    t_test_result <- t.test(formula, data = continuous_vars)

    # store tidy results of t-test in data frame
    t_test_result_tidy <- broom::tidy(t_test_result)
    t_test_result_tidy$Variable <- var
    t_test_results <- rbind(t_test_results, t_test_result_tidy)

}

# clean up t test results and add mean difference
t_test_results <- t_test_results %>%
    select(Variable, estimate1, estimate2, p.value) %>%
    left_join(difference_df)

# create table
results_table <- kable(t_test_results, format = "latex", col.names = c("Variable",
    "Mean No Treatment", "Mean Treatment", "P-Value", "Mean Difference"),
    caption = "T-Test Results Summary") %>%
    kable_styling(font_size = 7, latex_options = "hold_position")

# display results table
results_table
```

Table 1: T-Test Results Summary

Variable	Mean No Treatment	Mean Treatment	P-Value	Mean Difference
IND_SR	57.38515	48.55968	0	-8.83
COMM_VAL	117.22839	84.87908	0	-32.35

- **The mean difference in species richness index between the treated (regulated by ITQ) and untreated groups is -8.83, and the mean difference in commercial value between the two groups is -32.35. This means that on average, the treated group has a 8.83 lower species richness index than the untreated group. Also, on average, the treated group has a 32.35 lower commercial value than the untreated group. The t-test results show that these differences are statistically significant (using a significance level of alpha = 0.05). For both covariates, the p-value is equal to approximately 0. These results provide evidence that there are differences in pre-treatment charactaristics between the treated**

and untreated groups, and it is possible that these covariates may be influencing the treatment effects.

```
## Covariate Imbalance pre matching:
matching_balance <- xBalance(ITQ ~ MET + IND_SR + COMM_VAL, data = fisheries_data_new,
    report = c("std.diffs", "chisquare.test", "p.values"))

# Balance Table
balance_table <- kable(matching_balance, format = "latex", caption = "Covariate Imbalance before matchi
    kable_styling(font_size = 7, latex_options = "hold_position")

# Displaying the table
balance_table
```

Table 2: Covariate Imbalance before matching

	std.diff.unstrat	p.unstrat
METMET2	-0.0088168	0.6426817
METMET3	-0.0522914	0.0059460
METMET4	0.0905561	0.0000019
METMET5	0.1917710	0.0000000
METMET6	-0.0984729	0.0000002
IND_SR	-0.7959157	0.0000000
COMM_VAL	-0.7674422	0.0000000

	chisquare	df	p.value
unstrat	3004.041	7	0

- **Additionally, based on the covariate imbalance table, the chi-square value is very high, so we would reject the assumption that pre-treatment characteristics are unrelated to being chosen into a certain treatment group.**

Question (c) Treatment Ignorability *(1 pt)*

(c) Based on your results from (a) and (b), do you see a problem with just comparing the outcome variable means between treated and untreated fisheries?

- **There is a problem comparing the just the outcome variable between the treated and untreated fisheries. This is because our tests show that there is an imbalance between pre-treatment characteristics, and treatment selection is not completely random. If the selection of treatment was truly random, there would be no statistical difference between the distributions of these covariates. This means that treatment may not be the only variable influencing the outcome variable, and these covariates may act as confounding variables that influence the outcome variable as well. We cannot assume that *only* the treatment variable is influencing treatment effects.**

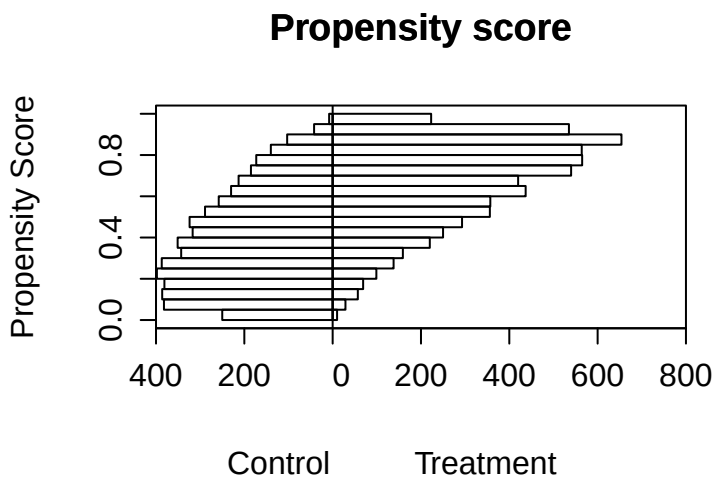
Question (d) Propensity Scores *(2 pts)*

(d) Estimate the propensity scores (probability of being treated) using a logit model, assume that all covariates are relevant and should be included in the estimation [0.5 pt]. Draw separate histograms (back to back) of the propensity scores for the treated and the untreated group [0.5 pt]. Comment on the overlap, do you have any concerns? Why/why not? [1]

```r
## Propensity Score Estimates Propensity Scores
ps <- glm(ITQ ~ MET + IND_SR + COMM_VAL, data = fisheries_data_new,
    family = binomial())

# use gtsummary to make table
library(gtsummary)
# print the table in a nice format
ps %>%
    tbl_regression()
```

Characteristic	**log(OR)**	**95% CI**	**p-value**
MET			
MET1	—	—	
MET2	0.10	-0.07, 0.27	0.2
MET3	0.00	-0.13, 0.13	>0.9
MET4	0.50	0.33, 0.67	<0.001
MET5	0.45	0.31, 0.59	<0.001
MET6	-0.06	-0.23, 0.11	0.5
IND_SR	-0.08	-0.09, -0.08	<0.001
COMM_VAL	-0.02	-0.02, -0.02	<0.001

```r
## PS Histogram
fisheries_data_new$psvalue <- predict(ps, type = "response")

histbackback(split(fisheries_data_new$psvalue, fisheries_data_new$ITQ),
    main = "Propensity score", xlab = c("Control", "Treatment"),
    ylab = "Propensity Score")
```



- **The propensity scores show the probability of individuals with certain characteristics to be selected for treatment. The overlap of propensity scores shows if the control group act as good counterfactuals for the treatment group by showing whether the two treatment groups (treatment/control groups) have the similar propensity scores. Looking at this histogram, it appears that while there is some overlap in propensity scores within the two treatment groups, the distribution of propensity scores are still different between the**

two groups. There are more treated individuals at higher propensity scores, but there very few control counterfactuals to choose from at these propensity scores. This means that it will be difficult to find good counterfactuals for most of these treated individuals. Additionally, there are not many treated individuals with lower propensity scores, but there are many more control individuals at these values.

Question (e) ATT with Nearest Neighbor Matching *(3 pts: 2 pt estimate, 1 pt interpretation)*

(e) Use the propensity scores from (c) to estimate the Average Treatment Effect on the Treated (ATT) with a nearest neighbor matching estimator. Interpret the result (just the size of the estimate)

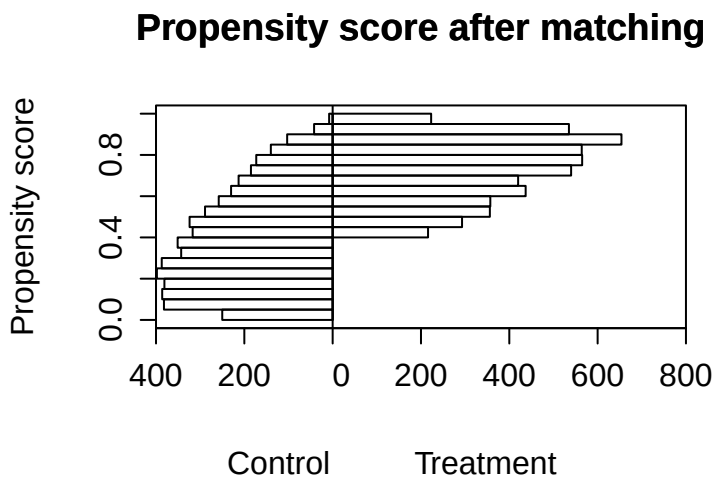
```
## Nearest Neighbor Matching
m.nn <- matchit(ITQ ~ MET + IND_SR + COMM_VAL, data = fisheries_data_new,
    method = "nearest", ratio = 1)
match.data = match.data(m.nn)

## Covariate Imbalance post matching:
matching_balance <- xBalance(ITQ ~ MET + IND_SR + COMM_VAL, data = match.data,
    report = c("std.diffs", "chisquare.test", "p.values"))

# Balance Table
matching_balance_table <- kable(matching_balance, format = "latex",
    caption = "Covariate Imbalance after matching") %>%
    kable_styling(font_size = 7, latex_options = "hold_position")

# Displaying the table
matching_balance_table
```

Table 3: Covariate Imbalance after matching

	std.diff.unstrat	p.unstrat
METMET2	-0.0100960	0.6080677
METMET3	-0.0679507	0.0005613
METMET4	0.1113937	0.0000000
METMET5	0.2288324	0.0000000
METMET6	-0.1150712	0.0000000
IND_SR	-0.9923428	0.0000000
COMM_VAL	-0.9411097	0.0000000

	chisquare	df	p.value
unstrat	3955.989	7	0

```
## Histogram of PS after matching
histbackback(split(match.data$psvalue, match.data$ITQ), main = "Propensity score after matching",
    xlab = c("Control", "Treatment"), ylab = "Propensity score")
```



- **Based on the propensity scores after matching and the covariate imbalance table after matching, it actually appears that there is a higher imbalance after nearest-neighbor matching. However, for the purposes of this assignment I will leave the matching method as nearest-neighbor. For real world applications, I assume that finding a different method of matching would work better to find better counterfactuals for the treatment group. The results in the rest of this section will likely be somewhat inaccurate due to the lack of good counterfactuals.**

```
# calculate ATT
diff_data <- match.data %>%
    group_by(subclass) %>%
    summarise(diff = mean(COLL_SHARE[ITQ == 1]) - mean(COLL_SHARE[ITQ ==
        0]), .groups = "drop")

ATT <- mean(diff_data$diff, na.rm = TRUE)
att_match_table <- kable(ATT, format = "latex", caption = "ATT after matching") %>%
    kable_styling(font_size = 7, latex_options = "hold_position")

att_match_table
```

Table 4: ATT after matching

x
-0.0713262

- **The average treatment effect after matching is -0.0713262. This means that the share of years a fishery is collapsed between 1990 and 2012 averages 0.0713262 less when treated (regulated with an ITQ) than when not treated (not regulated with ITQ).**

Question (f) ATE with WLS *(3 pts: 1 pt estimate, 1 pt interpretation)*

(f) Estimate the Average Treatment Effect (ATE) using the weighted least squares on the full sample. Interpret the estimated size and conclude if it is significantly different from zero from a statistical perspective.

```r
## WLS Matching prep
PS <- fisheries_data_new$psvalue
Y <- fisheries_data_new$COLL_SHARE
D <- fisheries_data_new$ITQ

# calculate weight
fisheries_data_new$wgt = (D/PS + (1 - D)/(1 - PS))

# linear model to understand ATE with WLS
reg_wls_c <- lm(COLL_SHARE ~ ITQ + MET + IND_SR + COMM_VAL, data = fisheries_data_new,
    weights = wgt)
# summary(reg_wls_c)

# present results
wls_table <- kable(summary(reg_wls_c)$coefficients, format = "latex",
    caption = "ATE with WLS matching") %>%
    kable_styling(font_size = 7, latex_options = "hold_position")
wls_table
```

Table 5: ATE with WLS matching

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.2628933	0.0027751	94.732804	0
ITQ	-0.0766783	0.0009902	-77.435631	0
METMET2	-0.0166478	0.0019198	-8.671735	0
METMET3	0.0150585	0.0014403	10.455407	0
METMET4	0.0484681	0.0019530	24.817606	0
METMET5	0.0796048	0.0015797	50.390854	0
METMET6	0.1119518	0.0019244	58.173963	0
IND_SR	-0.0014749	0.0000412	-35.804328	0
COMM_VAL	0.0003786	0.0000112	33.944026	0

- **The ATE using WLS is -0.0766783. This estimate tells us that the share of years a fishery is collapsed between 1990 and 2012 averages 0.0766783 less among the fisheries regulated with an ITQ than those that are not. The p-value is approximately zero, so using a significance level of alpha = 0.05, the average treatment effect is significantly different than 0, meaning that the effect of being regulated with an ITQ is statistically significant.**

2 Part 2 Difference in Difference Estimation *(10 points total + 3pts extra credit)*

Here we return for a final time to the dataset from Gertler, Martinez, and Rubio-Codina (2012) and use a different way of estimating the effect of the Mexican conditional cash transfer on the value of animal holdings of recipients. We'll use the panel data from assignment 2, where you have both the pre-program and post-program observations. See Template for dataset preparation instructions.

Data Preparation

*Note: You will need to install the packages plm and dplyr (included in template preamble). Again, you can find a description of the variables at the bottom of PDF and HERE.

Prepare Data: Load the new data (progresa_pre_1997.csv) and the follow-up data (progresa_post_1999.csv) into R. Note that we created a time denoting variable (with the same name, 'year') in BOTH datasets. Again, you will create a panel dataset by appending the data (i.e. binding the dataset row-wise together

creating a single dataset). We want to examine the same outcome variable as before, value of family animal holdings (vani). You will use the full dataset for each estimate. NOTE: you should not change any NAs from the TREATED column in your analysis, as we expect that spillover was likely in this program. NAs will be excluded from your calculations/estimations.

Question (a) DiD Estimator, ATE *(5 pts: 3 pts estimate, 2 pts interpretation)*

(a) Calculate the DiD estimator of the treatment effect (ATE) of the program on the value of animal holdings (vani) "manually" i.e. based on group mean values without running a regression. Report and interpret the result (Note: no significance test or standard errors is possible, so you do not need to report these values).

```
## Estimate ATE with DiD estimator manually.  You will need
## to calculate various means to get this estimate

# separate data to make means easier ---- treated after and
# before ---- treated 1999
YT_after_data <- progresa_full %>%
    filter(treatment == 1, year == 1999)
# treated 1997
YT_before_data <- progresa_full %>%
    filter(treatment == 1, year == 1997)

# untreated after and before ---- untreated 1999
YC_after_data <- progresa_full %>%
    filter(treatment == 0, year == 1999)
# untreated 1997
YC_before_data <- progresa_full %>%
    filter(treatment == 0, year == 1997)

# calculate means ---- treatment means
YT_after_mean <- mean(YT_after_data$vani)
YT_before_mean <- mean(YT_before_data$vani)
delta_YT <- YT_after_mean - YT_before_mean

# control means
YC_after_mean <- mean(YC_after_data$vani)
YC_before_mean <- mean(YC_before_data$vani)
delta_YC <- YC_after_mean - YC_before_mean

## Compute the Difference-in-Differences ---- treatment -
## control
did <- round((delta_YT) - (delta_YC), 2)
did
```

```
## [1] 287.9
```

- **When calculated manually, the difference in difference estimator on the ATE of value of animal holdings is $287.90. This means that after the treatment time period, households that received treatment had an average of $287.90 higher value of animal holdings than households that did not receive treatment. This estimate allows us to understand the effect of treatment, taking into account that there may be time-invariant unobserved characteristics. DID helps us to interpret treatment effect disregarding other time-invariant characteristics that also change the outcome variable over time.**

- **When looking at the mean value of animal holdings over time for each group separately, it appears that there is a decrease in value after the treatment time period for both groups. The control group had a decrease in value of animal holdings by an average of $1156.75 over time, while the treatment group had an average decrease in value of animal holdings by $868.85. The DID shows us that although both groups decreased in value of animal holdings over time, the treatment may have allowed less of a decrease for treated households, as their post-treatment value is, on average, still $287.90 higher than the control group.**

Question (b) Difference in Difference using OLS *(5 pts)*

(b) Now set up an OLS-regression using group mean values to estimate the same ATE. Interpret the estimated treatment effect [3 pts]. Also interpret the coefficients on the time dummy and the group dummy variable (see interpretation done in class in lecture 9) [2 pts].

Hints: You will need to create a new dataframe with a variety of dummy variables to do this. The R example provided with the DiD module (and/or the excel file) should help.

```
## Create a new data frame for OLS regression
progresa_ols <- progresa_full %>%
    mutate(post_treat = ifelse(year == 1999, 1, 0), treatment = ifelse(treatment ==
        1, 1, 0), time_x_treated = (post_treat * treatment)) %>%
    group_by(post_treat, treatment, time_x_treated) %>%
    summarise(vani = mean(vani))

## Run the OLS regression w/dummies
model_did <- lm(vani ~ treatment + post_treat + time_x_treated,
    data = progresa_ols)

## Report OLS Model results Print the summary of the OLS
## model
summary(model_did)
```

```
##
## Call:
## lm(formula = vani ~ treatment + post_treat + time_x_treated,
##     data = progresa_ols)
##
## Residuals:
## ALL 4 residuals are 0: no residual degrees of freedom!
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2848.2        NaN     NaN      NaN
## treatment        -237.7        NaN     NaN      NaN
## post_treat      -1156.8        NaN     NaN      NaN
## time_x_treated    287.9        NaN     NaN      NaN
##
## Residual standard error: NaN on 0 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:      1,  Adjusted R-squared:    NaN
## F-statistic:   NaN on 3 and 0 DF,  p-value: NA
```

```
# create table
ols_did_table <- kable(summary(model_did)$coefficients, format = "latex",
    caption = "DID with OLS (group means)") %>%
    kable_styling(font_size = 7, latex_options = "hold_position")

ols_did_table
```

Table 6: DID with OLS (group means)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2848.2175	NaN	NaN	NaN
treatment	-237.6927	NaN	NaN	NaN
post_treat	-1156.7517	NaN	NaN	NaN
time_x_treated	287.9050	NaN	NaN	NaN

- **The DID using OLS with group means is $287.90. This means that using the OLS method, the value of animal holdings after treatment in treated households is, on average, $287.90 higher than untreated households. The treatment group coefficient shows us that before treatment, the treatment group's value of animal holdings was an average of $237.70 lower than the untreated group. The time dummy coefficient shows us that the value of animal holdings among the untreated group decreased by an average of $1156.80 over the treatment time period.**

3 Extra Credit: ATE with OLS using full dataset *(3 pts: 2 pts estimate, 1 pt interpretation)*

(c) Estimate the ATE with an OLS-regression based on the original units as observations (i.e. not with group mean values, you will need to use the entire dataset). Even though the specification is the same as in the regression with the group mean values above, you'll need to create new indicator variables for the treatment group and the post treatment time period as well as their interaction term. Verify that you get the same result as above. Now report also on the precision of the estimation and test whether the estimated coefficient is different from zero.

```
## Create the dummy variables (you'll need 3)
progresa_ols_ec <- progresa_full %>%
    mutate(post_treat = ifelse(year == 1999, 1, 0), time_x_treated = (post_treat *
        treatment))

## OLS regression
model_did_ec <- lm(vani ~ treatment + post_treat + time_x_treated,
    data = progresa_ols_ec)

# summary
summary(model_did_ec)
```

```
##
## Call:
## lm(formula = vani ~ treatment + post_treat + time_x_treated,
##     data = progresa_ols_ec)
##
```

```
## Residuals:
##    Min     1Q Median     3Q    Max
##  -2848  -1953  -1495    -62  81456
##
## Coefficients:
##                Estimate Std. Error t value             Pr(>|t|)
## (Intercept)     2848.22      60.61  46.989 < 0.0000000000000002 ***
## treatment       -237.69      80.45  -2.954              0.00314 **
## post_treat     -1156.75      85.72 -13.494 < 0.0000000000000002 ***
## time_x_treated   287.90     113.78   2.530              0.01140 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4633 on 27024 degrees of freedom
##   (1724 observations deleted due to missingness)
## Multiple R-squared:  0.01169,    Adjusted R-squared:  0.01158
## F-statistic: 106.6 on 3 and 27024 DF,  p-value: < 0.00000000000000022
```

```r
# Present Regressions in Table
ols_did_table_ec <- kable(summary(model_did_ec)$coefficients,
    format = "latex", caption = "DID with OLS") %>%
    kable_styling(font_size = 7, latex_options = "hold_position")

ols_did_table_ec
```

Table 7: DID with OLS

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2848.2175	60.61455	46.989010	0.0000000
treatment	-237.6927	80.45311	-2.954425	0.0031352
post_treat	-1156.7517	85.72191	-13.494235	0.0000000
time_x_treated	287.9050	113.77788	2.530413	0.0113985

- **Using OLS with the complete dataset results in the same values as using the group means in the previous question. The DID is still 287.90, and the coefficients are also the same (just rounded slightly differently). The DID estimate has a p-value of 0.01140 and is statistically significant at a significance level of alpha = 0.05. This means that the DID estimate is significantly different than 0.**