UTRGV

TREATMENT EFFECT ESTIMATION USING INVERSE PROBABILITY TREATMENT WEIGHTING

Presented by Gaukhar Nurbek

Main Topics

Points to talk about

Objectives

Theory

Methods

Results and simulation

Chalenges

The Project's Objectives

TO STUDY HOW IPTW IS ESTIMATED

TO STUDY METHODS THAT ARE USED IN TREATMENT EFFECT ESTIMATION

TO WRITE FUNCTION WHICH CALCULATES IPTW

Theory

Treatment effect

The term ‘treatment effect’ refers to the causal effect of a binary (0–1) variable on an outcome variable of scientific or policy interest.

The term ‘treatment effect’ originates in a medical literature concerned with the causal effects of binary, yes-or-no ‘treatments’, such as an experimental drug or a new surgical procedure.

Treatment effects can be estimated using social experiments, regression models, matching estimators, and instrumental variables.

Presented by Gaukhar Nurbek

Theory

Propensity score

is defined as the probability of treatment assignment conditional on measured baseline covariates.

Propensity score methods are being used with increasing frequency to estimate treatment effects using observational data.

$e = P(Z = 1|X)$

Theory

Inverse probability of treatment weighting

is the one of the method of using propensity score.

$w = Z/e + (1-Z)/(1-e)$

Each subject's weight is equal to the inverse of the probability of receiving the treatment that the subject received.

METHODS OF USING PROPENSITY SCORE TO ESTIMATE TREATMENT EFFECT

1

Weighting

The first set of propensity-score estimators use the propensity scores as weights to create a balanced sample of treated and control observations.

$$\tilde{\tau} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{Z_i Y_i}{e(X_i)} - \frac{(1-Z_i)Y_i}{1-e(X_i)}\right)$$

2

Conditional expectation

estimate the conditional expectation of Y given W and e(X). Treatment effect can be estimated by

$$v_w(e) = \mathbb{E}[\mathrm{Y}(\mathrm{Z})|e(X) = e]$$

$$\widetilde{\tau_{reg}} = \frac{1}{N}\sum_{i=1}^{N}(\widehat{v_1}(e(X_i)) - \widehat{v_0}(e(X_i)))$$

3

Weighting + Regression

One can rewrite the weighting estimator as the estimating following regression function by weighted least squares

$$Y_i = \alpha + \tau * W_i + \varepsilon_i$$

BUILDING FUNCTIONS TO ESTIMATE TREATMENT EFFECT

GLR (x,y,w)

Linear regression model function which accepts treatment, output and weights as an input
Returns the estmated coeficients, which is a treatment effect estimation

IPTW(treatment, output, covs, data)

Inverse probability treatment weighting function accepts as a parameters treatment column name, output column name, co-varities column name and data
Output: data frame with a new column like ps, ps_recieved, and IPTW weights.

BUILDING FUNCTIONS TO ESTIMATE TREATMENT EFFECT

````
#### IPTW calculator function

```{r}
IPTW = function(treatment,output,covs,data)
{
#ps and iptw estimation
  ps  = data  %>%
  group_by_(covs) %>%
  summarise(ps = n()/sum(data[treatment]))

data_ps  = data %>%
  #this merges the propensity score into the bigger dataset by levels
  left_join(ps, by = c(covs)) %>%
  #if exposed, assign the propensity score. if unexposed assign 1-the
  mutate(ps_received = case_when(
    data[treatment]==1 ~ ps,
    data[treatment]==0 ~ 1-ps)) %>%
  #and the weights (for IPTW) are 1 over the ps_received
  mutate(IPTW = weekend/ps_received + (1-weekend)/(1-ps_received))

return(data_ps)
}
```
````

````
#### Linear regression model functinon

```{r}
GLR = function(x,y,weights){
one = rep(1,length(y))
weight = sqrt(weights)
x = x*weight
y = y*weight
X = cbind(one,x)
b = solve(t(X) %*% X) %*% t(X) %*% y
J = matrix(rep(1,length(y)*length(y)),ncol = length(y),nrow = length(y))
SSTO = t(y) %*% y - 1/(length(y)) * t(y) %*% J %*% y
SSE = t(y) %*% y - t(b) %*% t(X) %*% y
SSR = t(b) %*% t(X) %*% y - ((1/length(y))*t(y)) %*% J %*% y
MSR = SSR/(length(b) - 1)
MSE = SSE/(length(y) - length(b))
F_S = MSR/MSE
R_S = SSR/SSTO
res = y - X%*%b
predicted = X%*%b
residual_st_error = sqrt(sum(res^2)/(length(y) - length(b)))
resid_results = summary(res)
b_hat = solve(t(X) %*% X)
SE = sqrt(diag(b_hat))*residual_st_error
tb1 = b/SE
pvalue1 = 2 - 2*pt(abs(tb1), df = length(y)- length(b))
coefficient_results = cbind(b=b,SE=SE,tb1=tb1,pvalue1=pvalue1)
p_value_F = pf(F_S, length(b) - 1, length(y) - length(b), lower.tail = FALSE)
newList = list("coefficients" = coefficient_results, "residuals" = resid_results ,df = length(y) - length(b),
residual_st_error, "F*" = F_S, "R^2" = R_S, "p_value" = p_value_F)
return(list(newList=newList,res=res,predicted=predicted))
}
```
````

Simulation

Dataset: Movies domestic gross income from Jun-Dec 2018

Real data from **https://www.boxofficemojo.com/**

was partially changed to get the weekend column(which is a treatment)

Date	Day	DayNu...	Top10.Gross	YD	LW	Releases	Film	Gross	GrossNew
<int>	<int>	<int>	<fctr>	<fctr>	<fctr>	<int>	<fctr>	<fctr>	<fctr>
43465	1	365	$36,240,441	-19.60%	-14.20%	53	Aquaman	$10,011,638	$10011638
43464	7	364	$50,932,176	-12.40%	2.90%	51	Aquaman	$16,440,551	$16440551
43463	6	363	$58,118,460	2.60%	4.60%	51	Aquaman	$18,632,907	$18632907
43462	5	362	$56,667,767	9.70%	-2.90%	51	Aquaman	$17,041,113	$17041113
43461	4	361	$51,671,321	-7%	299.20%	53	Aquaman	$14,622,228	$14622228
43460	3	360	$55,579,761	-21.60%	261.80%	52	Aquaman	$16,903,518	$16903518
43459	2	359	$70,898,743	-3.40%	30%	53	Aquaman	$21,982,419	$21982419
43458	1	358	$31,752,644	-35.80%	184.50%	49	Aquaman	$10,851,928	$10851928
43457	7	357	$49,473,471	-10.90%	72.70%	53	Aquaman	$18,793,733	$18793733
43456	6	356	$55,554,266	-4.80%	29.80%	52	Aquaman	$21,346,634	$21346634

Goal: To estimate effect of the weekend on the gross revenue

If it is weekday treatment = 0 if its weekend treatment = 1

Covariates: Film

Output: Revenue per day

Presented by Gaukhar Nurbek

Treatment effect estimation

Results

````
```{r}
iptw_data = IPTW('weekend','GrossRevenue','Film',new_df)
outDatWeight = data.frame(outcome = iptw_data$GrossRevenue, treatment = iptw_data$weekend, wt = iptw_data$IPTW)
x = outDatWeight$treatment
y = outDatWeight$outcome
w = outDatWeight$wt
head(iptw_data)
fit = GLR(x, y, w)
fit$newList$coefficients
```
````

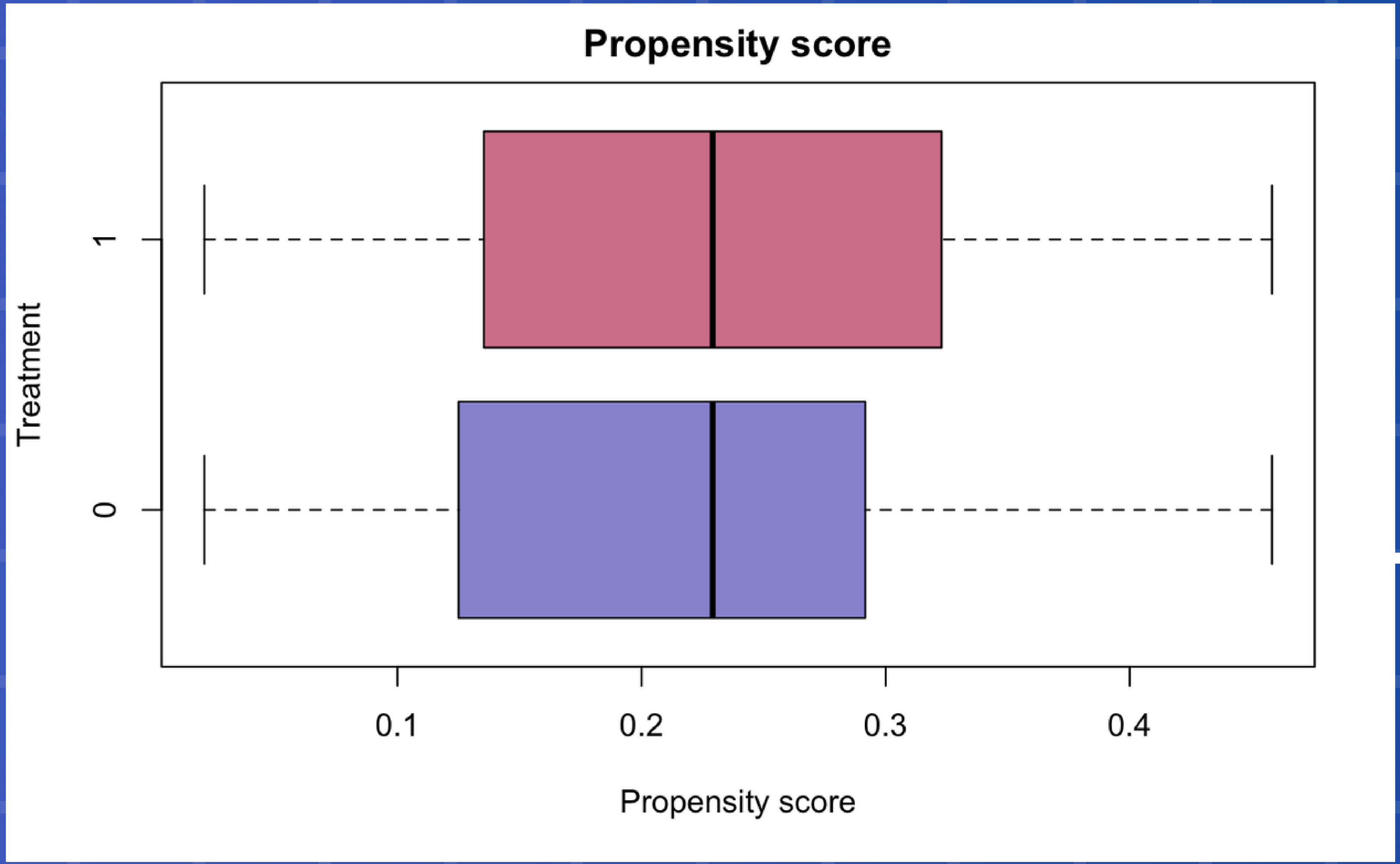
data.frame
6 x 15

R Console

LW	Releases	Film	Gross	GrossNew	GrossRevenue	weekend	ps	ps_received	IPTW
<fctr>	<int>	<fctr>	<fctr>	<fctr>	<int>	<dbl>	<dbl>	<dbl>	<dbl>
-14.20%	53	Aquaman	$10,011,638	$10011638	10011638	0	0.2291667	0.7708333	4.363636
2.90%	51	Aquaman	$16,440,551	$16440551	16440551	1	0.2291667	0.2291667	4.363636
4.60%	51	Aquaman	$18,632,907	$18632907	18632907	1	0.2291667	0.2291667	4.363636
-2.90%	51	Aquaman	$17,041,113	$17041113	17041113	0	0.2291667	0.7708333	4.363636
299.20%	53	Aquaman	$14,622,228	$14622228	14622228	0	0.2291667	0.7708333	4.363636
261.80%	52	Aquaman	$16,903,518	$16903518	16903518	0	0.2291667	0.7708333	4.363636

```
                          SE
one 15465369 1341927.8 11.524740 0.000000e+00
x    7425855  959891.3  7.736141 9.654499e-13
```

Results



Propensity scores

Chalenges

FORMATTING INPUT PARAMETERS FOR IPTW FUNCTION

IPTW IS USED TO CALCULATE ONLY AVERAGE TREATMENT EFFECT

TREATMENT EFFECT ESTIMATION METHODS NEED TO BE STUDIED MORE

Thank you!
Any questions?

Presented by Gaukhar Nurbek