

Lab: Propensity Score Weighting with Multiple Treatments

Fan Li

Contents

0.1	Setup: Loading Packages	1
1	Multiple treatment groups example: synthetic MEPS data	1
1.1	The synthetic MEPS Data Set	1
1.2	Design: Propensity Score Estimation and Balance Check	3
1.3	Analysis: Weighting and Augmented Weighting	6

0.1 Setup: Loading Packages

```
library(PSweight)
```

1 Multiple treatment groups example: synthetic MEPS data

This document is about showing how to apply the PSweight package to study multiple treatment groups using balancing weights method. The illustration will be based on a Synthetic version of the MEPS data set. This analysis will be focusing on unconfounded descriptive comparisons.

1.1 The synthetic MEPS Data Set

We use a synthetic data set of the 2009 Medical Expenditure Panel Survey (MEPS) data. Our synthetic data set includes 20 categorical covariates and 5 continuous covariates. The primary outcome of interest is the total health expenditure for each individual in 2009 (healthExp). We wish to study the health disparities between Asian, Black, White, and Hispanic (race).

We will analyze this synthetic data set using several different approaches including inverse probability weighting, overlap weighting, and weighting method target at the average treatment effect for the treated (ATT) - defined based on a specific reference group. For each approach, we are interested in estimating six marginal (marginal means over the specific target population) contrasts of two out of four groups, i.e., the average controlled difference between two groups regarding the medical expenditure. We report both the point estimates, the robust sandwich standard errors, and the corresponding 95% confidence interval.

```
#Load the synthetic MEPS Data Set  
meps <- read.csv("meps_syn.csv")  
#Transfer race to categorical variables  
meps$race <- factor(meps$race)
```

Below table summarizes the interpretations of most of the covariates (Li and Li, 2019).

Name	Interpretations
pcs	SF-12 physical component summary
mcs	SF-12 mental component summary
age	age
bmi	body mass index
sinceCheckUp	time since last general checkup
male	gender
health	self-reported physical health status (five categories)
mhealth	self-reported mental health status (five categories)
anylim	any limitation of activity
socialLim	any limitation of social participation
cogLim	any limitation of cognitive functions
exercise	exercise
hibp	history of high blood pressure
chd	coronary heart disease
mhealth	emphysema
emphysema	emphysema
cholesterol	high cholesterol
cancer	cancer
stroke	stroke
angina	angina
arthritis	arthritis
asthma	asthma
mi	myocardial infarction
smoke	smoking status

```
#Present the summary statistics of the outcome and the race
summary(meps$healthExp)
```

```
  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
    0     122     984   4570   4006 253186
```

```
summary(meps$race)
```

```
  Asian   Black Hispanic   White
  1458   4003   5221   9764
```

```
#Ensure the type of categorical variables is correct
```

```
meps_rep <- meps
```

```
meps_rep[,-c(1:4,10,11,27)] <- lapply(meps_rep[,-c(1:4,10,11,27)], factor)
```

```
#Present the summary statistics of all covariates
```

```
summary(meps_rep[,-c(1,27)])
```

```
      pcs           mcs      sinceCheckUp  smoke      socialLim cogLim
Min.   : 7.12   Min.   : 1.45   Min.   :1.000   0:16801   0:19582   0:19602
1st Qu.:43.87   1st Qu.:44.69   1st Qu.:1.000   1: 3645   1:  864   1:  844
Median :53.03   Median :53.09   Median :1.000
Mean   :49.13   Mean   :50.62   Mean   :1.999
3rd Qu.:56.71   3rd Qu.:57.33   3rd Qu.:2.000
Max.   :71.10   Max.   :73.01   Max.   :6.000
exercise  male           age           bmi      married  health
0: 8874   0:11189   Min.   :18.00   Min.   : 9.40   0: 9655   1:4667
1:11572   1: 9257   1st Qu.:31.00   1st Qu.:23.70   1:10791   2:6503
           Median :44.00   Median :27.10           3:6209
```

```

                Mean    :45.63    Mean    :28.06                4:2374
                3rd Qu.:58.00    3rd Qu.:31.20                5: 693
                Max.    :85.00    Max.    :81.40
mhealth  hibp      chd      angina    mi          stroke    emphysema
1:7247   0:13645   0:19220   0:19909   0:19689   0:19732   0:20033
2:6162   1: 6801    1: 1226   1: 537    1: 757    1: 714    1: 413
3:5507
4:1286
5: 244

cholesterol cancer    diabetes arthritis asthma    anylim
0:14212   0:18632   0:18468   0:15520   0:18532   0:15116
1: 6234   1: 1814   1: 1978   1: 4926   1: 1914   1: 5330

```

1.2 Design: Propensity Score Estimation and Balance Check

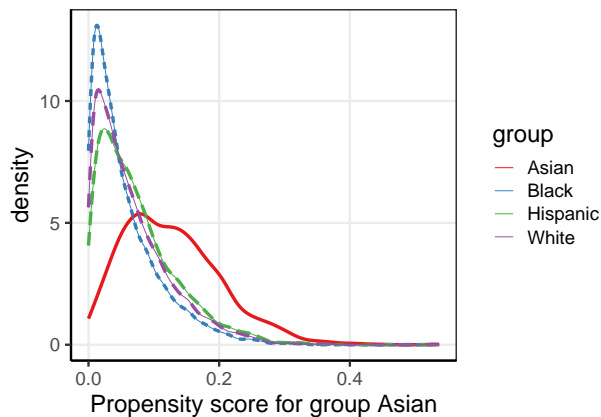
Firstly, we fit the propensity score model by a multinomial logistic regression and visually check the overlap of the data.

```

#Specify a multinomial logistic regression model: ps.meps
ps.meps <- race ~ pcs + mcs + sinceCheckUp + as.factor(smoke) +
  as.factor(socialLim) + as.factor(cogLim) + as.factor(exercise) +
  as.factor(male) + age + bmi + as.factor(married) + as.factor(health) +
  as.factor(mhealth) + as.factor(hibp) + as.factor(chd) +
  as.factor(angina) + as.factor(mi) + as.factor(stroke) + as.factor(emphysema) +
  as.factor(cholesterol) + as.factor(cancer) + as.factor(diabetes) +
  as.factor(arthritis) + as.factor(asthma) + as.factor(anylim)
#Obtain the propensity score estimates
bal.meps <- SumStat(ps.formula = ps.meps, weight = c('IPW', 'overlap', 'treated'),
  data = meps)
#Density plots of estimated propensity scores
plot(bal.meps, type = 'density')

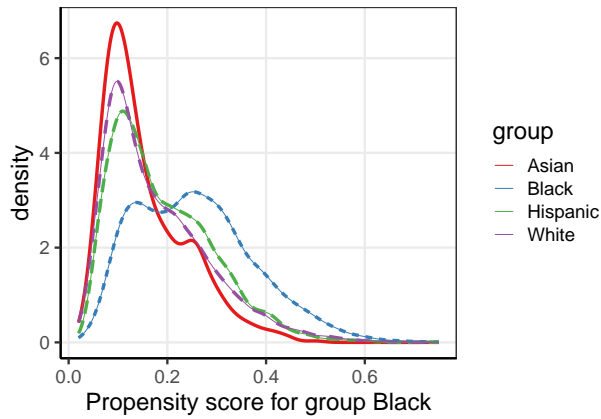
```

Propensity score for group Asian



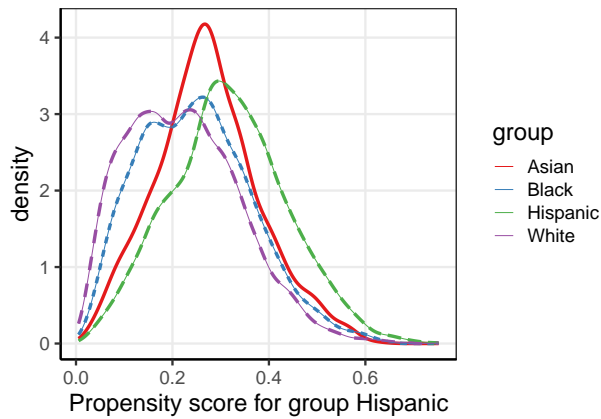
Press [enter] to continue

Propensity score for group Black



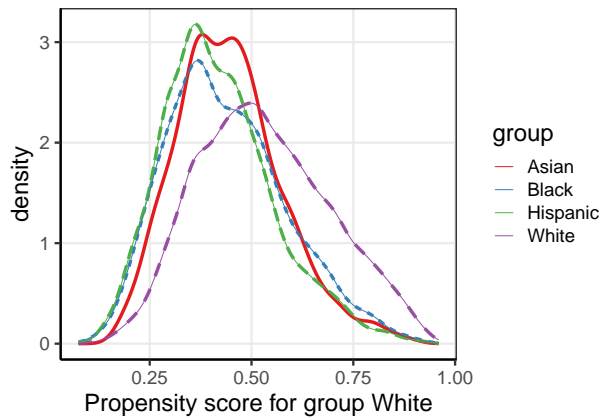
Press [enter] to continue

Propensity score for group Hispanic



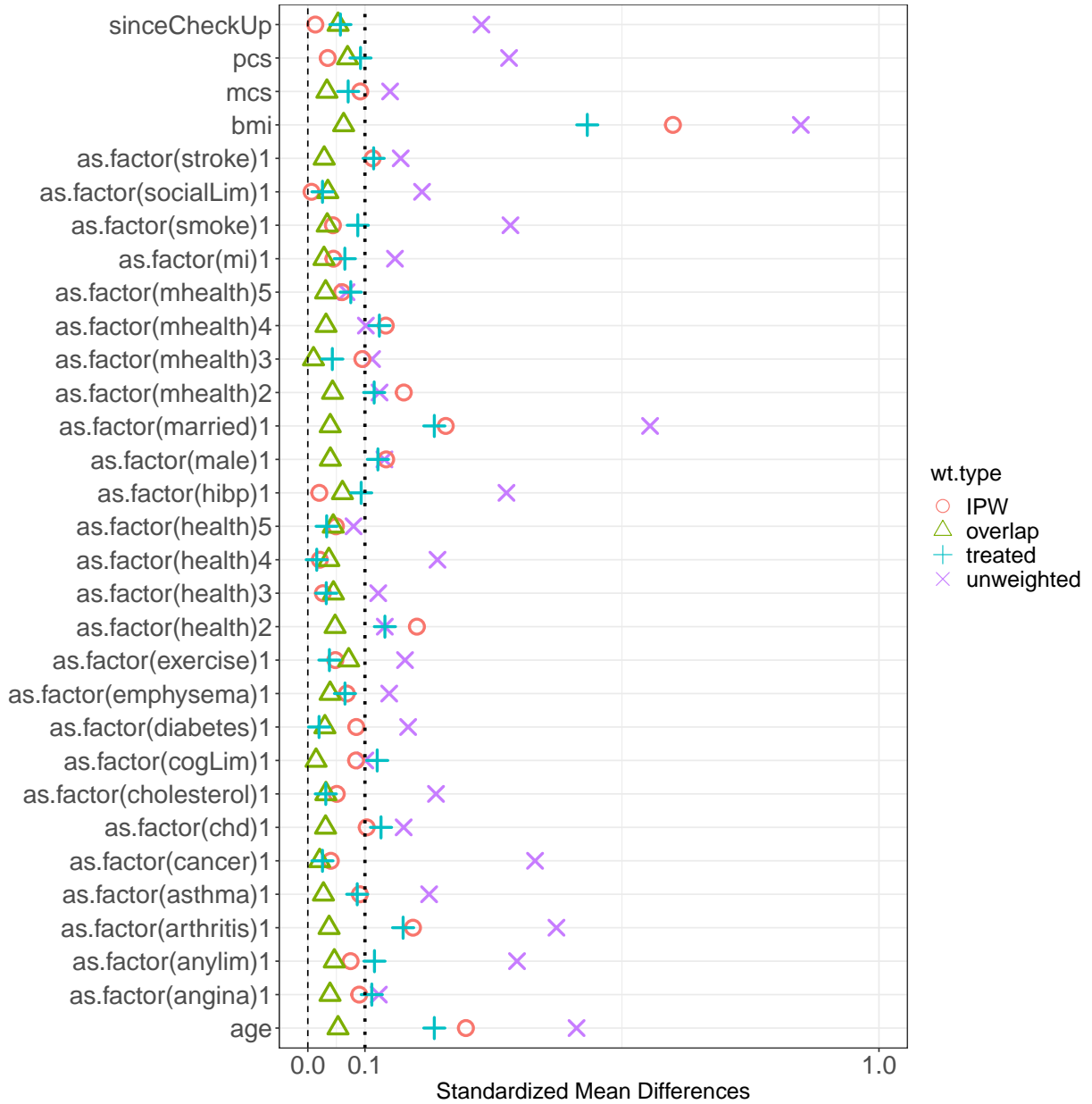
Press [enter] to continue

Propensity score for group White



We can see sufficient overlap of the estimated generalized propensity score for each group although the estimated generalized propensity scores for Asian and Black are right-skewed. There are also units with very small generalized propensity score to belong to the Asian group.

```
#Check covariates balance using the metric average standardized differences (ASD)
plot(bal.meps, metric = 'ASD')
```



From the love plot, we can conclude that OW method balances all covariates relatively well (every covariates have smaller standardized mean difference). However, IPW and ATT weighting methods no longer provide an effective improvement of covariates balance compared to the binary treatment case, as some of covariates have a larger standardized mean difference than the unweighted.

1.3 Analysis: Weighting and Augmented Weighting

1.3.1 Weighting

We use IPW method first and compare the standard errors estimated by robust sandwich estimator (default choice) or bootstrap.

```
#Implement IPW weighting method and estimating standard errors by robust sandwich
#estimator
ate.meps.ipw <- PSweight(ps.formula = ps.meps, yname = "healthExp", data = meps,
                        weight = 'IPW')
#Implement IPW weighting method and estimating standard errors by bootstrap
set.seed(0)
ate.meps.ipw.boot <- PSweight(ps.formula = ps.meps, yname = "healthExp",
                              data = meps, weight = 'IPW', bootstrap = TRUE,
                              R=50)
```

bootstrap 50 samples

```
#Target estimands are the contrasts between any two of races
contrasts.mult <- rbind(c(-1,1,0,0),c(-1,0,1,0),c(-1,0,0,1),c(0,-1,1,0),
                       c(0,-1,0,1),c(0,0,-1,1))
#Estimates
summary(ate.meps.ipw, type = 'DIF',
        constrast=contrasts.mult)
```

Closed-form inference:

Original group value: Asian, Black, Hispanic, White

Contrast:

	Asian	Black	Hispanic	White
Contrast 1	-1	1	0	0
Contrast 2	-1	0	1	0
Contrast 3	-1	0	0	1
Contrast 4	0	-1	1	0
Contrast 5	0	-1	0	1
Contrast 6	0	0	-1	1

	Estimate	Std.Error	lwr	upr	Pr(> z)	
Contrast 1	1763.884	426.250	928.449	2599.32	3.501e-05	***
Contrast 2	1683.969	525.208	654.581	2713.36	0.001345	**
Contrast 3	2272.107	367.263	1552.284	2991.93	6.148e-10	***
Contrast 4	-79.915	460.401	-982.284	822.45	0.862198	
Contrast 5	508.223	266.463	-14.036	1030.48	0.056483	.
Contrast 6	588.138	405.432	-206.495	1382.77	0.146879	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
summary(ate.meps.ipw.boot, type = 'DIF',
        constrast=contrasts.mult)
```

Use Bootstrap sample for inference:

Original group value: Asian, Black, Hispanic, White

Contrast:

	Asian	Black	Hispanic	White
Contrast 1	-1	1	0	0
Contrast 2	-1	0	1	0
Contrast 3	-1	0	0	1
Contrast 4	0	-1	1	0
Contrast 5	0	-1	0	1
Contrast 6	0	0	-1	1

	Estimate	Std.Error	lwr	upr	Pr(> z)
Contrast 1	1763.884	469.202	730.415	2404.24	<2e-16 ***
Contrast 2	1683.969	460.716	578.043	2417.75	<2e-16 ***
Contrast 3	2272.107	405.019	1335.323	2832.25	<2e-16 ***
Contrast 4	-79.915	423.886	-825.774	953.37	0.82
Contrast 5	508.223	239.198	53.100	956.79	0.06 .
Contrast 6	588.138	338.776	-65.936	1192.17	0.12

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
summary(ate.meps.ipw, type = 'DIF',
        constrast=contrasts.mult)$estimates[,c(1,2,4,5)]
```

	Estimate	Std.Error	lwr	upr
Contrast 1	1763.88396	426.2503	928.44868	2599.3192
Contrast 2	1683.96883	525.2075	654.58101	2713.3567
Contrast 3	2272.10661	367.2632	1552.28398	2991.9292
Contrast 4	-79.91513	460.4006	-982.28375	822.4535
Contrast 5	508.22265	266.4632	-14.03555	1030.4809
Contrast 6	588.13778	405.4323	-206.49490	1382.7705

```
summary(ate.meps.ipw.boot, type = 'DIF',
        constrast=contrasts.mult)$estimates[,c(1,2,4,5)]
```

	Estimate	Std.Error	lwr	upr
Contrast 1	1763.88396	469.2018	730.41462	2404.2398
Contrast 2	1683.96883	460.7160	578.04299	2417.7485
Contrast 3	2272.10661	405.0191	1335.32320	2832.2458
Contrast 4	-79.91513	423.8863	-825.77365	953.3747
Contrast 5	508.22265	239.1982	53.09958	956.7909
Contrast 6	588.13778	338.7759	-65.93630	1192.1724

The estimates of standard error by robust sandwich estimator and bootstrap are similar qualitatively, but still have quite some difference. Here for illustration we set the number of bootstrap replicate to be 20, which is likely not adequate. We should try at least 100, but it takes much longer to run. Hereafter, we only present the robust sandwich estimator of the standard error.

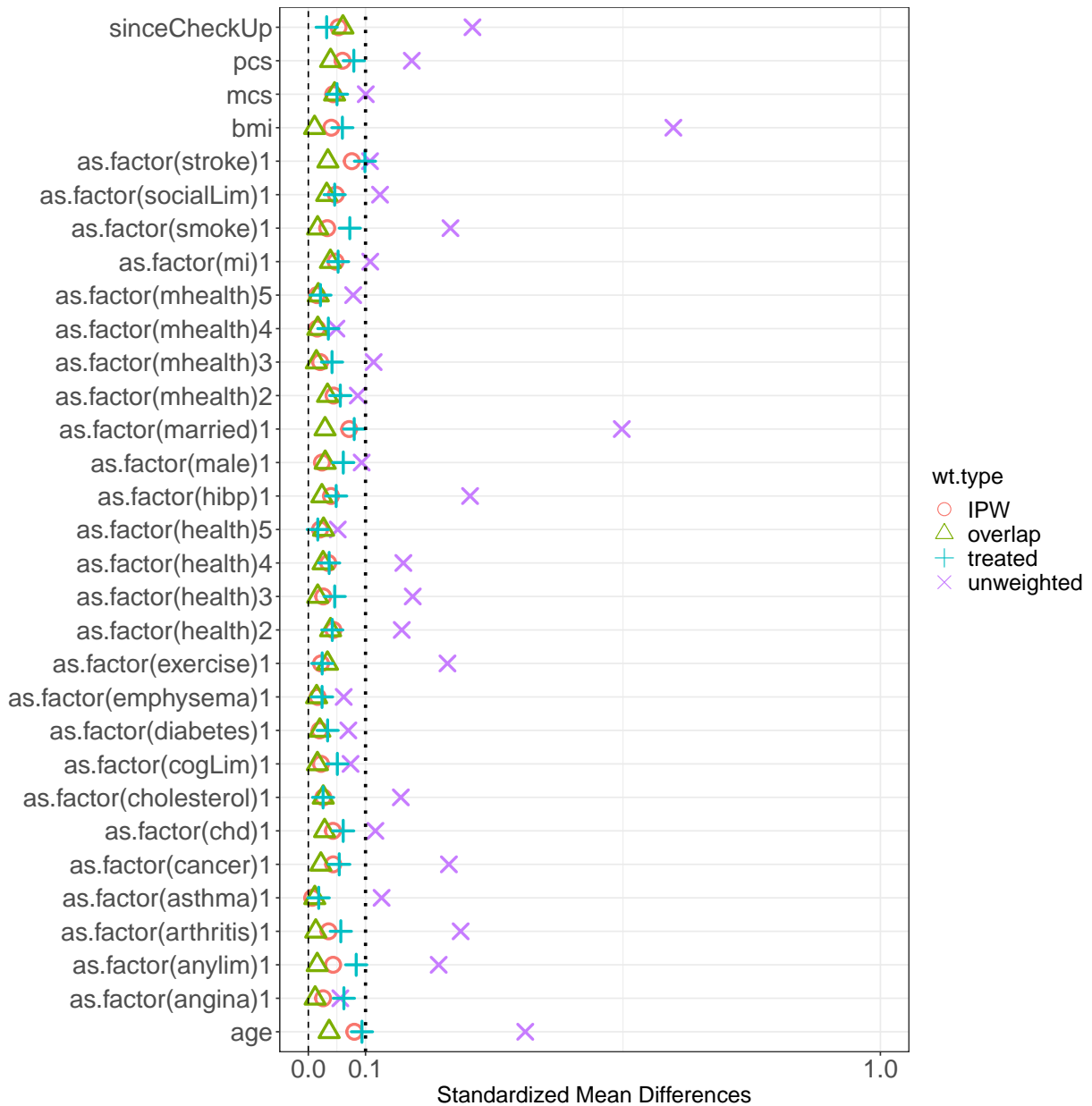
The point estimates of six contrasts between any two of four racial groups using IPW weighting method are respectively 1763.9, 1684.0, 2272.1, -79.9, 508.2, 588.1. The corresponding standard error estimates are 426.3, 525.2, 367.3, 460.4, 266.5, 405.4. The 95% confidence intervals are [928.4, 2599.3], [654.6, 2713.4], [1552.3, 2991.9], [-982.3, 822.5], [-14.0, 1030.5], [-206.5, 1382.8] respectively.

Next, we consider IPW trimming method to avoid undue influence of units with extremely small propensity scores. Firstly, We use the fixed threshold trimming with the threshold $\delta = 0.03$, i.e., excluding those whose smallest generalized propensity scores are less than 0.03.

```

#Apply trimming with delta=0.03
bal.meps.trim <- SumStat(ps.formula = ps.meps,
  weight = c('IPW','overlap','treated'),
  data = meps, delta = 0.03)
#Covariates balance after trimming
plot(bal.meps.trim, metric = 'ASD')

```



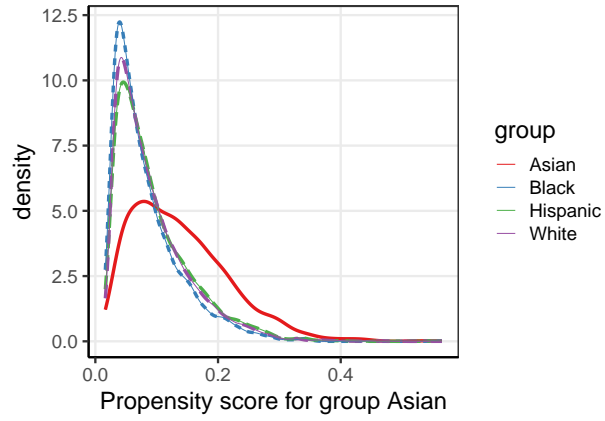
We can see that the covariates balance after trimming is preserved and very similar to that without trimming.

```

#Check the overlap
#Density plots of estimated propensity scores after trimming
plot(bal.meps.trim, type = 'density')

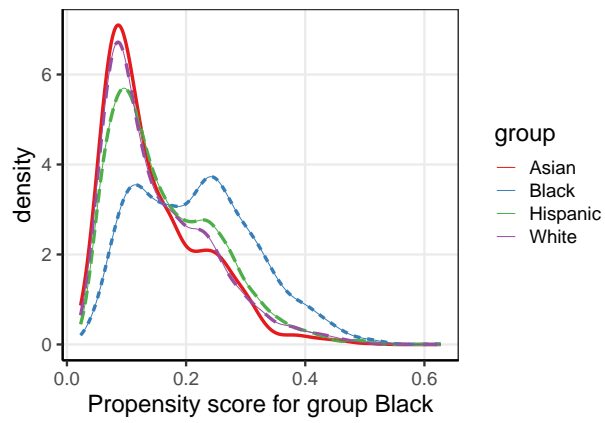
```

Propensity score for group Asian



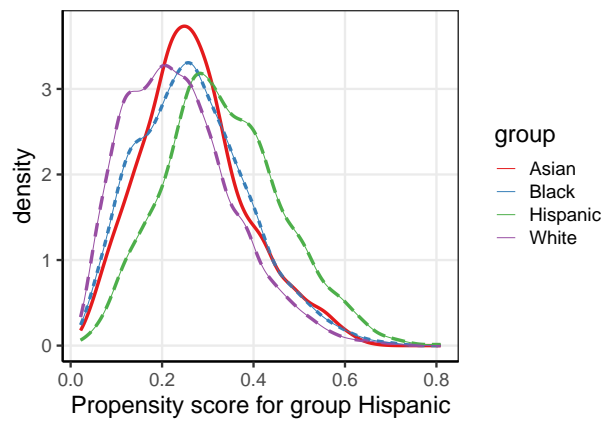
Press [enter] to continue

Propensity score for group Black



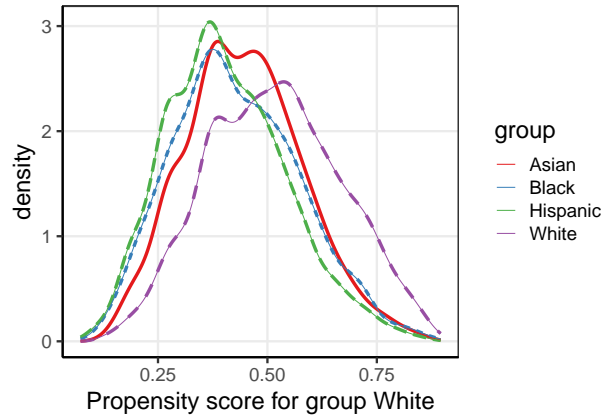
Press [enter] to continue

Propensity score for group Hispanic



Press [enter] to continue

Propensity score for group White



Overlap is still maintained as expected.

```
#Check how many patients are excluded
bal.meps.trim$trim
```

```

      Asian Black Hispanic White
trimmed   99  1636      1357  3162
remained 1359  2367      3864  6602
```

However, we can see that nearly 30.6% patients are trimmed and they are mostly from the Black, Hispanic and White racial groups. We then estimate the causal risk difference using IPW after trimming.

```
#Implement IPW weighting method after trimming
ate.meps.ipw.trim <- PSweight(ps.formula = ps.meps, yname = "healthExp",
                             data = meps,
                             weight = 'IPW',delta=0.03)

#Estimates
summary(ate.meps.ipw.trim,
        type = 'DIF', contrast=contrast.rhc)$estimates[,c(1,2,4,5)]
```

```

      Estimate Std.Error      lwr      upr
Contrast 1  486.67355  349.0672 -197.485609 1170.8327
Contrast 2  529.09068  296.3387  -51.722570 1109.9039
Contrast 3 1100.73497  213.8669  681.563478 1519.9065
Contrast 4   42.41713  376.7049 -695.910857  780.7451
Contrast 5  614.06142  317.5562  -8.337337 1236.4602
Contrast 6  571.64429  256.3241   69.258339 1074.0302
```

The point estimates using IPW weighting method after trimming are respectively 486.7, 529.1, 1100.7, 42.4, 614.1, 571.6. The corresponding standard error estimates are respectively 349.1, 296.3, 213.9, 376.7, 317.6, 256.3. The 95% confidence intervals are $[-197.5, 1170.8]$, $[-51.7, 1109.9]$, $[681.6, 1519.9]$, $[-695.9, 780.7]$, $[-8.3, 1236.5]$, $[69.3, 1974.0]$.

Next, we consider to implement the optimal trimming procedure provided by PSweight package. The optimal trimming procedure minimizes the total asymptotic variance of all pairwise contrasts within the class of symmetric trimming rules.

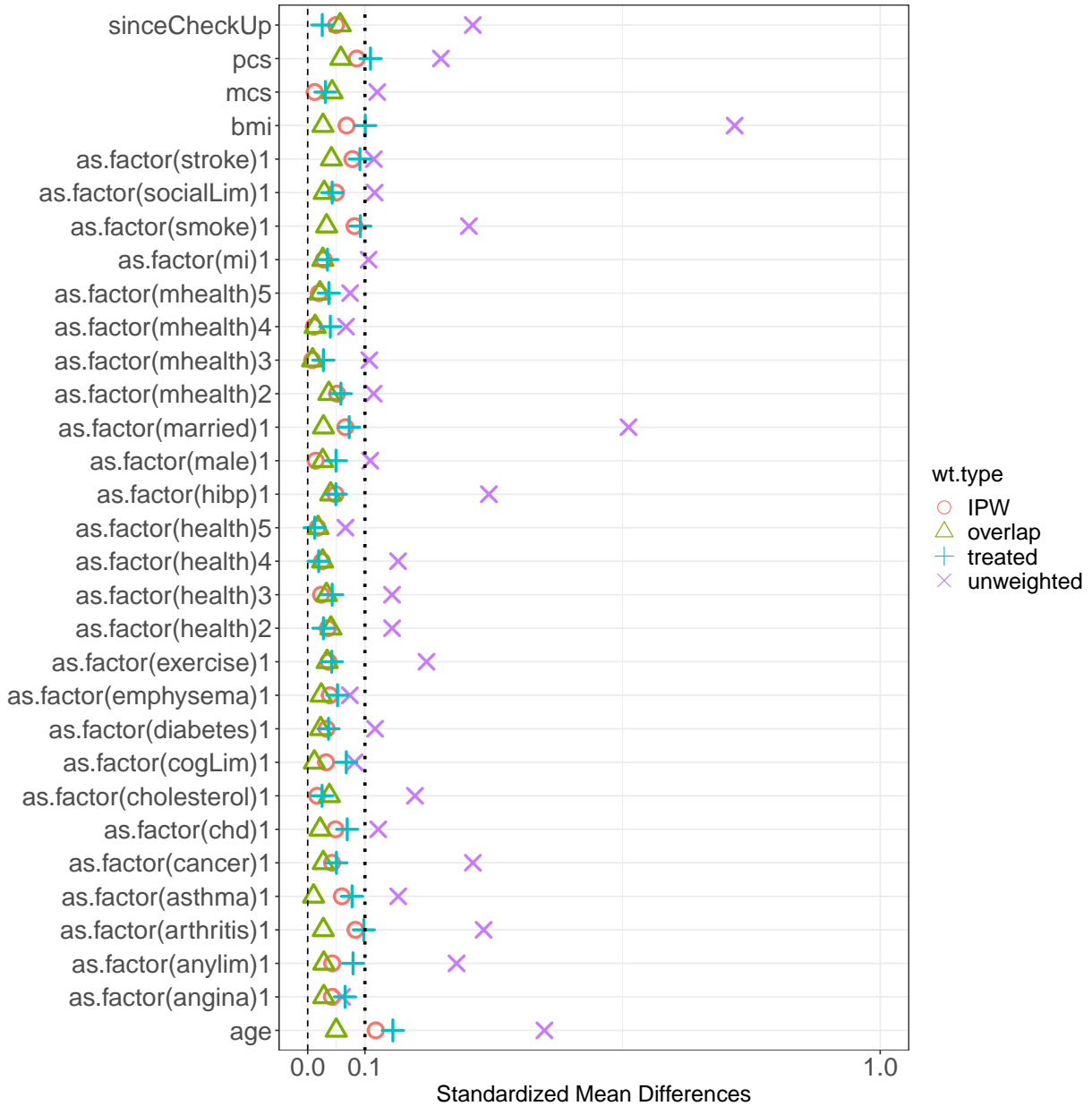
```
#Optimal trimming
opt_trim <- PStrim(ps.formula = ps.meps, data = meps, optimal = TRUE)
#Check how many patients are trimmed
opt_trim$trim_sum
```

```

      Asian Black Hispanic White
trimmed   49   917      602  1761
remained 1409  3086      4619  8003
```

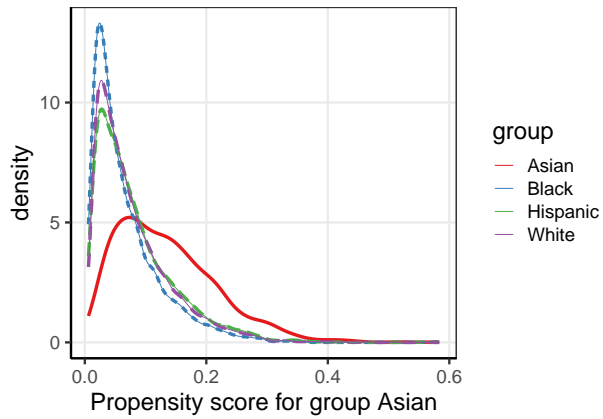
We can see that much fewer patients are trimmed using optimal trimming procedure compared to the above, fixed threshold trimming.

```
#Extract the trimmed data
opt_trim_meps <- opt_trim$data
#Check the overlap
bal.meps.trim.opt <- SumStat(ps.formula = ps.meps,
  weight = c('IPW', 'overlap', 'treated'),
  data = opt_trim_meps)
#Covariates balance after trimming
plot(bal.meps.trim.opt, metric = 'ASD')
```



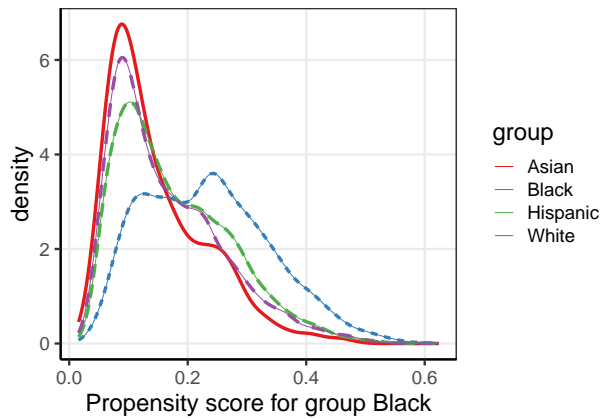
```
#Check the overlap
#Density plots of estimated propensity scores after trimming
plot(bal.meps.trim.opt, type = 'density')
```

Propensity score for group Asian



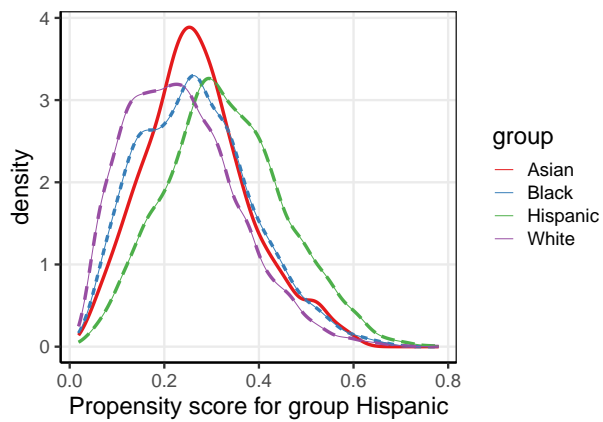
Press [enter] to continue

Propensity score for group Black



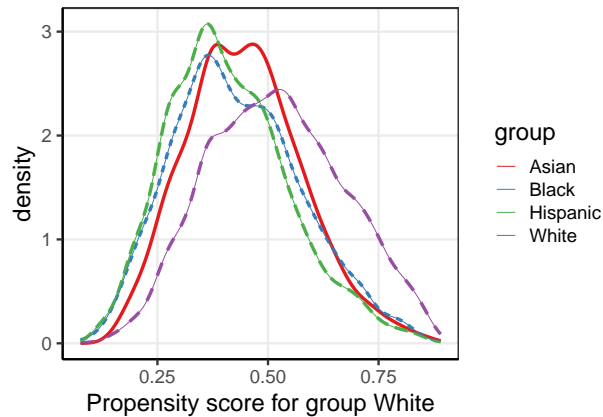
Press [enter] to continue

Propensity score for group Hispanic



Press [enter] to continue

Propensity score for group White



```
#Implement IPW weighting method after optimal trimming
ate.meps.ipw.trim.opt <- PSweight(ps.formula = ps.meps, yname = "healthExp",
                                data = opt_trim_meps,
                                weight = 'IPW')

#Estimates
summary(ate.meps.ipw.trim.opt,
        type = 'DIF', contrast=contrast.rhc)$estimates[,c(1,2,4,5)]
```

	Estimate	Std.Error	lwr	upr
Contrast 1	472.7739	332.8845	-179.66773	1125.2156
Contrast 2	580.9783	321.4221	-48.99739	1210.9541
Contrast 3	1227.8875	261.6480	715.06693	1740.7081
Contrast 4	108.2044	312.8007	-504.87364	721.2825
Contrast 5	755.1136	252.1868	260.83661	1249.3905
Contrast 6	646.9091	234.9059	186.50202	1107.3163

The point estimates using IPW method after optimal trimming are respectively 472.8, 581.0, 1227.9, 108.2, 755.1, 646.9. The corresponding standard error estimates are respectively 332.9, 321.4, 261.6, 312.8, 252.2, 234.9. The 95% confidence intervals are respectively [-179.7, 1125.2], [-49.0, 1211.0], [715.1, 1740.7], [-504.9, 721.3], [260.8, 1249.4], [186.5, 1107.3].

Next, we use ATT weighting method. Because we do not specify the “treated” reference population, this would be based on the last value in alphabetic order - White.

```
#Implement ATT weighting method
ate.meps.att <- PSweight(ps.formula = ps.meps, yname = "healthExp", data = meps,
                        weight = 'treated')

#Estimates
summary(ate.meps.att, type = 'DIF',
        contrast=contrasts.mult)$estimates[,c(1,2,4,5)]
```

	Estimate	Std.Error	lwr	upr
Contrast 1	1868.3989	519.4885	850.2201	2886.578
Contrast 2	1974.5567	772.8915	459.7173	3489.396
Contrast 3	2127.5682	419.2562	1305.8411	2949.295
Contrast 4	106.1578	737.1359	-1338.6021	1550.918
Contrast 5	259.1693	348.2863	-423.4594	941.798
Contrast 6	153.0114	669.2296	-1158.6545	1464.677

The point estimates of six contrasts between any two of four races using ATT weighting method are respectively 1868.4, 1974.6, 2127.6, 106.2, 259.2, 153.0. The corresponding standard error estimates are 519.5, 772.9, 419.3, 737.1, 348.3, 669.2 respectively. The 95% confidence interval are [850.2, 2886.6], [459.7, 3489.4], [1305.8, 2949.3], [-1338.6, 1550.9], [-423.5, 941.8], [-1158.7, 1464.7] respectively.

Next, we use OW weighting method.

```
#Implement overlap weighting (OW) method
ate.meps.ow <- PSweight(ps.formula = ps.meps, yname = "healthExp", data = meps,
                       weight = 'overlap')

#Estimates
summary(ate.meps.ow, type = 'DIF',
        constrast=contrasts.mult)$estimates[,c(1,2,4,5)]
```

	Estimate	Std.Error	lwr	upr
Contrast 1	573.04004	271.3975	41.11072	1104.9694
Contrast 2	653.77892	273.5201	117.68930	1189.8685
Contrast 3	1140.11540	194.9113	758.09619	1522.1346
Contrast 4	80.73888	299.8417	-506.94008	668.4178
Contrast 5	567.07536	234.0014	108.44100	1025.7097
Contrast 6	486.33648	235.5683	24.63104	948.0419

The point estimates of six contrasts between any two of four races using OW method are respectively 573.0, 653.8, 1140.1, 80.7, 567.1, 486.3. The corresponding standard error estimates are 271.4, 273.5, 194.9, 299.8, 234.0, 235.6 respectively. The 95% confidence interval are [41.1, 1105.0], [117.7, 1189.9], [758.1, 1522.1], [-506.9, 668.4], [108.4, 1025.7], [24.6, 948.0] respectively.

Overall, we can see that OW method generally has the smallest standard errors or the narrowest confidence intervals.

1.3.2 Augmented Weighting

In this section, we introduce the augmented weighting method and follow the same procedure as the previous section.

Using the augmented IPW method:

```
#Propose a logistic regression model for the outcome: out.meps
out.meps <- healthExp ~ pcs + mcs + sinceCheckUp + as.factor(smoke) +
  as.factor(socialLim) + as.factor(cogLim) + as.factor(exercise) +
  as.factor(male) + age + bmi + as.factor(married) + as.factor(health) +
  as.factor(mhealth) + as.factor(hibp) + as.factor(chd) +
  as.factor(angina) + as.factor(mi) + as.factor(stroke) + as.factor(emphysema) +
  as.factor(cholesterol) + as.factor(cancer) + as.factor(diabetes) +
  as.factor(arthritis) + as.factor(asthma) + as.factor(anylim)

#Implement augmented IPW weighting method
ate.meps.ipw.aug <- PSweight(ps.formula = ps.meps, yname = "healthExp",
                             data = meps,
                             out.formula = out.meps, weight = 'IPW',
                             augmentation = TRUE)

#Estimates
summary(ate.meps.ipw.aug, type = 'DIF',
        constrast=contrasts.mult)$estimates[,c(1,2,4,5)]
```

	Estimate	Std.Error	lwr	upr
Contrast 1	1364.1376	404.8331	570.67929	2157.5960
Contrast 2	1238.0165	486.4954	284.50296	2191.5300
Contrast 3	1939.5283	345.1535	1263.03992	2616.0168
Contrast 4	-126.1212	429.8559	-968.62329	716.3809
Contrast 5	575.3907	260.0400	65.72173	1085.0597
Contrast 6	701.5119	374.8389	-33.15878	1436.1825

The point estimates of six contrasts between any two of four races using the augmented IPW method are respectively 1364.1, 1238.0, 1939.5, -126.1, 575.4, 701.5. The corresponding standard error estimates are 404.8, 486.5, 345.2, 429.9, 260.0, 374.8 respectively. The 95% confidence intervals are [570.7, 2157.6], [284.5, 2191.5], [1263.0, 2616.0], [-968.6, 716.4], [65.7, 1085.1], [-33.2, 1436.2] respectively.

Using the augmented ATT method (again with the White group as the reference group):

```
#Implement augmented ATT weighting method
set.seed(0)
ate.meps.att.aug <- PSweight(ps.formula = ps.meps, yname = "healthExp",
                             data = meps,
                             out.formula = out.meps, weight = 'treated',
                             augmentation = TRUE, bootstrap = TRUE, R=20)

#Estimates
summary(ate.meps.att.aug, type = 'DIF',
        contrast=contrasts.mult)$estimates[,c(1,2,4,5)]
```

The point estimates of six contrasts between any two of four races using the augmented ATT weighting method are respectively 1767.7, 1708.3, 2132.9, -59.3, 365.2, 424.6. The corresponding standard error estimates are 448.9, 891.4, 492.9, 777.1, 297.0, 647.3 respectively. The 95% confidence intervals are [1234.1, 2601.0], [682.9, 3550.9], [1590.0, 3178.1], [-1041.0, 1542.4], [98.8, 1046.2], [-757.3, 1326.4] respectively.

Using the augmented OW method:

```
#Implement augmented IPW weighting method
ate.meps.ow.aug <- PSweight(ps.formula = ps.meps, yname = "healthExp",
                             data = meps,
                             out.formula = out.meps, weight = 'overlap',
                             augmentation = TRUE)

#Estimates
summary(ate.meps.ow.aug, type = 'DIF',
        contrast=contrasts.mult)$estimates[,c(1,2,4,5)]
```

	Estimate	Std.Error	lwr	upr
Contrast 1	349.0002	262.2000	-164.90241	862.9028
Contrast 2	495.3807	259.7314	-13.68347	1004.4448
Contrast 3	1201.3525	189.0085	830.90262	1571.8023
Contrast 4	146.3805	286.7626	-415.66397	708.4249
Contrast 5	852.3523	227.0940	407.25618	1297.4484
Contrast 6	705.9718	223.1736	268.55959	1143.3840

The point estimates of six contrasts between any two of four races treatments using the augmented OW method are respectively 349.0, 495.4, 1201.4, 146.4, 852.4, 706.0. The corresponding standard error estimates are 262.2, 259.7, 189.0, 286.8, 227.1, 223.2 respectively. The 95% confidence interval are [-164.9, 862.9], [-13.7, 1004.4], [830.9, 1571.8], [-415.7, 708.4], [407.3, 1297.4], [268.6, 1143.4] respectively.

Overall, we can see that OW method generally has the smallest standard errors and thus the narrowest confidence intervals.