

# Homework2

Jon Campbell

```
library(tidyverse)
library(knitr)
library(PSweight)
```

## Part 1

### Question 1

```
hsr <- read.table("data/hw2/HSR.txt", header = TRUE)
#p_z = P(Y(z)==1)
#target estimand is ATE p_1 - p_2
X <- apply(hsr[,c(1,3,4,5,6,7,8,12)], MARGIN = 2, factor)
Z <- factor(hsr$pg)
Y <- hsr$i_aqoc
```

a)

```
ps_formula <- as.factor(pg) ~ i_age + as.factor(i_sex) + as.factor(i_race) + as.factor(i_e

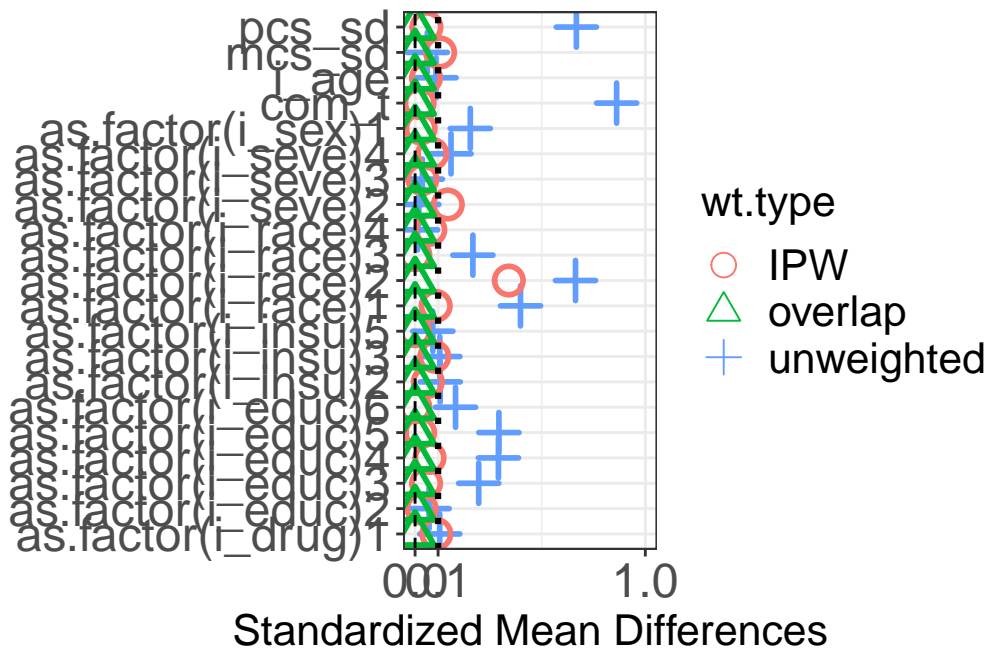
prop_scores <- glm(ps_formula, family = "binomial"
                  , data = hsr)$fitted.values

prop_score_summary <- SumStat(ps_formula = ps_formula
                             , weight = c('IPW','overlap')
                             ,data = hsr)

#Estimated propensity Scores
cat(prop_scores)
```

0.3666097 0.4691487 0.5608349 0.8235156 1.73677e-07 0.2859149 0.6462016 0.841273 0.3948395 0

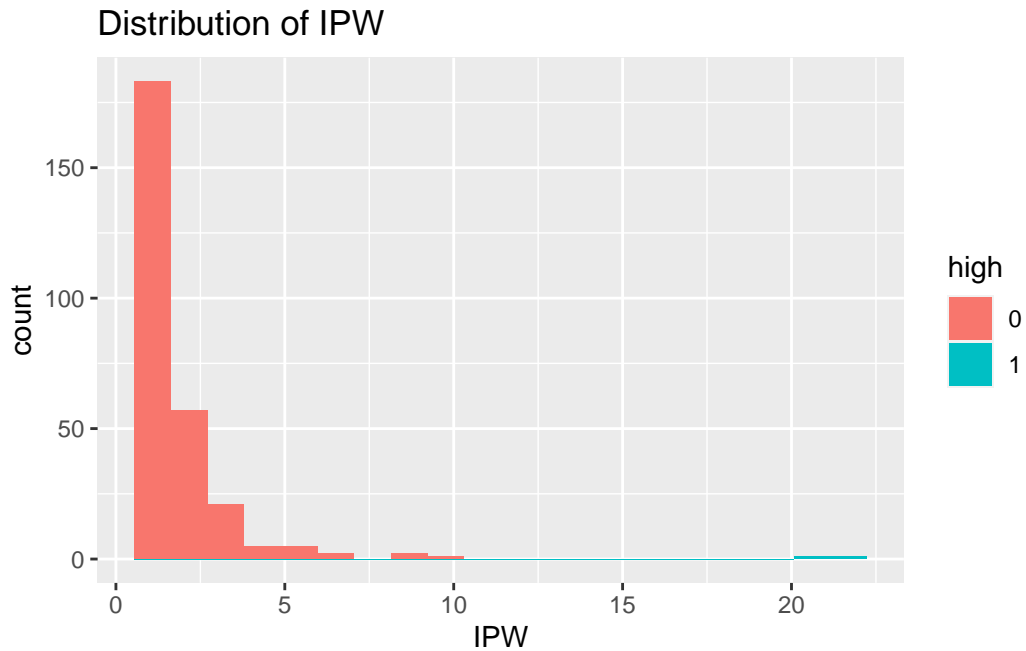
```
#checking balance via ASD
plot(prop_score_summary, metric = "ASD")
```



From the love plot we see that most of the covariates have an ASD above 0.1 when unweighted but perform much better with IPW. mcs\_sd and severity2 are slightly above 0.1 and race2 is around 0.35 under IPW. The covariates are perfectly balanced with overlap weighing.

```
#Inverse Probability Weights
IPW_weights <- ifelse(Z == 2, 1/prop_scores, 1/(1-prop_scores))

#Normalized IPW
IPW_weights_norm <- numeric(length(IPW_weights))
IPW_weights_norm[Z==2] = IPW_weights[Z==2]/sum(IPW_weights[Z==2])
IPW_weights_norm[Z==1] = IPW_weights[Z==1]/sum(IPW_weights[Z==1])
ggplot(data.frame(weights = IPW_weights_norm,
                  high=as.factor(ifelse(IPW_weights_norm>15,1,0)))
, aes(x = weights, fill = high)) +
  geom_histogram(bins = 20) +
  labs(title = "Distribution of IPW", x = "IPW")
```



Looking at the distribution of IPW there are 2 observations with high weights relative to the sample.

**b)**

```
tibble(  
  prop_score = prop_scores,  
  trt = Z  
) |>  
  ggplot(aes(x = prop_score, fill = trt)) +  
  geom_density(alpha = 0.5)
```



From the density plot of propensity scores by group, there is good overlap. The love plot shows good balance with overlap weighting and IPW weighting.

c)

```
unweighted <- prop_score_summary$unweighted.sumstat[,1:2]
colnames(unweighted) <- c("unweighted.trt1", "unweighted.trt2")

ipw_weighted <- prop_score_summary$IPW.sumstat[,1:2]
colnames(ipw_weighted) <- c("IPW.trt1", "IPW.trt2")

overlap_weighted <- prop_score_summary$overlap.sumstat[,1:2]
colnames(overlap_weighted) <- c("overlap.trt1", "overlap.trt2")

tab1 <- cbind(unweighted, ipw_weighted, overlap_weighted) |> round(4)
tab1 |>
  kable()
```

	unweighted.trt1	unweighted.trt2	IPW.trt1	IPW.trt2	overlap.trt1	overlap.trt2
i_age	40.4571	39.6532	41.5684	41.1479	41.1248	41.1248
as.factor(i_sex)1	0.7619	0.6532	0.7181	0.7071	0.7152	0.7152
as.factor(i_race)1	0.8381	0.6416	0.7601	0.7205	0.8236	0.8236
as.factor(i_race)2	0.0095	0.2197	0.0269	0.1371	0.0200	0.0200
as.factor(i_race)3	0.0571	0.0116	0.0374	0.0372	0.0380	0.0380
as.factor(i_race)4	0.0667	0.0694	0.0765	0.0597	0.0752	0.0752
as.factor(i_educ)2	0.0190	0.0116	0.0145	0.0114	0.0165	0.0165
as.factor(i_educ)3	0.0952	0.0289	0.0561	0.0455	0.0608	0.0608
as.factor(i_educ)4	0.3905	0.2254	0.2862	0.3152	0.3064	0.3064
as.factor(i_educ)5	0.2571	0.4277	0.3621	0.3514	0.3213	0.3213
as.factor(i_educ)6	0.2286	0.3064	0.2771	0.2765	0.2950	0.2950
as.factor(i_insu)2	0.2571	0.3064	0.2515	0.2761	0.2535	0.2535
as.factor(i_insu)3	0.0000	0.0058	0.0000	0.0035	0.0000	0.0000
as.factor(i_insu)5	0.0571	0.0405	0.0309	0.0307	0.0254	0.0254
as.factor(i_drug)1	0.9905	0.9769	0.9941	0.9846	0.9925	0.9925
as.factor(i_seve)2	0.2381	0.2312	0.2176	0.2794	0.2508	0.2508
as.factor(i_seve)3	0.4667	0.4509	0.3963	0.4118	0.4453	0.4453
as.factor(i_seve)4	0.1905	0.1329	0.1755	0.1478	0.1587	0.1587
com_t	2.9143	1.7168	2.4733	2.4398	2.3778	2.3778
pcs_sd	41.4148	48.2981	46.1089	46.6206	45.3764	45.3764
mcs_sd	48.9392	48.4198	48.1294	46.9954	48.4299	48.4299

d)

```
Z <- ifelse(Z == 1,0,1)
ps.hsr <- as.factor(pg) ~ i_age + as.factor(i_sex) + as.factor(i_race) + as.factor(i_educ)

ate.ipw.norm <- PSweight(ps.formula = ps.hsr, yname = "i_aqoc", data = hsr
, weight = "IPW")
ato.overlap <- PSweight(ps.formula = ps.hsr, yname = "i_aqoc", data = hsr
, weight = "overlap")
ate.ipw.nonnorm <- summary(lm(i_aqoc ~ pg, data = hsr, weights = IPW_weights))$coefficient
results.ate <- rbind(summary(ate.ipw.norm, type = "DIF", contrast = c(-1,1))$estimates[,1:2],
summary(ato.overlap, type = "DIF", contrast = c(-1,1))$estimates[,1:2],
ate.ipw.nonnorm)

rownames(results.ate) <- c("ATE IPW Normalized", "ATO Overlap", "ATE NonNormalized")
results.ate |> round(4)
```

	Estimate	Std.Error
ATE IPW Normalized	-0.1455	0.0562
ATO Overlap	-0.1272	0.0607
ATE NonNormalized	-0.1455	0.0527

## Question 2

a)

```
ps_formula <- as.factor(pg) ~ i_age + as.factor(i_sex) + as.factor(i_race) + as.factor(i_educ)

prop_scores <- glm(ps_formula, family = "binomial"
, data = hsr)$fitted.values

hsr_subsetted <- subset(hsr, i_sex == 1)
hsr_subsetted$ipw_weights <- ifelse(hsr_subsetted$pg == 1, 1/prop_scores, 1/(1-prop_scores))
att.subgroup <- summary(lm(i_aqoc ~ pg + i_age + i_sex + as.factor(i_race) + as.factor(i_educ), data = hsr_subsetted, weights = ipw_weights))$coefficient
att.subgroup
```

	Estimate	Std. Error
	-0.18540597	0.08938014

b)

```
ps_formula <- as.factor(pg) ~ i_age + as.factor(i_race) + as.factor(i_educ) + as.factor(i_ins)  
  
prop_scores <- glm(ps_formula, family = "binomial"  
  , data = hsr)$fitted.values  
  
hsr_subsetted <- subset(hsr, i_sex == 1)  
hsr_subsetted$ipw_weights <- ifelse(hsr_subsetted$pg == 1, 1/prop_scores, 1/(1-prop_scores))  
  
att.subgroup <- summary(lm(i_aqoc ~ pg + i_age + i_sex + as.factor(i_race) + as.factor(i_educ)  
  att.subgroup
```

Estimate	Std. Error
-0.18272986	0.08904017

c)

```
hsr_subsetted <- subset(hsr, i_sex == 1)  
ps.hsr <- as.factor(pg) ~ i_age + as.factor(i_race) + as.factor(i_educ) + as.factor(i_ins)  
  
summary(PWeight(ps_formula = ps.hsr, yname = "i_aqoc", data = hsr_subsetted  
  , weight = "IPW"))$estimates[,1:2]
```

Estimate	Std. Error
-0.13034248	0.06778558

### Question 3

a)

b)

c)

d)

e)

## Part 2

```
brscn <- read.table("data/hw2/brscn.txt", header = TRUE)
```

## Question 1