

Propensity Score Diagnostics

Lucy D'Agostino McGowan
Wake Forest University

2021-09-01

Checking balance

- Love plots (Standardized Mean Difference)
- ECDF plots

Standardized Mean Difference (SMD)

$$d = \frac{X_{\text{treatment}} - X_{\text{control}}}{\sqrt{\frac{s_{\text{treatment}}^2 + s_{\text{control}}^2}{2}}}$$

SMD in R

Calculate standardized mean differences

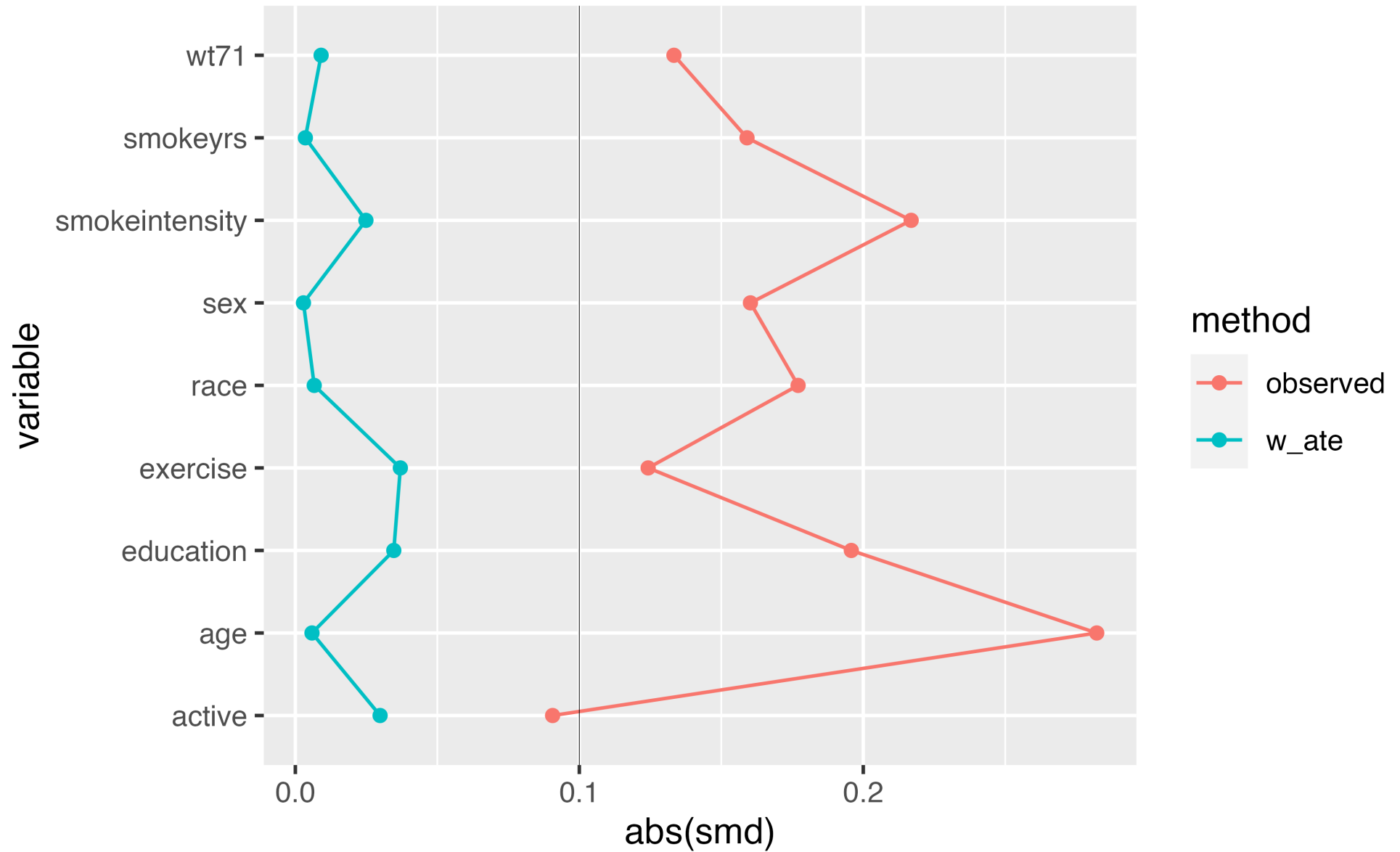
```
1 library(halfmoon)
2 library(tidyverse)
3
4 smds <- tidy_smd(
5   df,
6   .vars = c(confounder_1, confounder_2, ...),
7   .group = exposure,
8   .wts = wts # weight is optional
9 )
```

SMD in R

Plot them! (in a Love plot!)

```
1 ggplot(  
2   data = smds,  
3   aes(x = abs(smd), y = variable, group = weights, color = weights)  
4 ) +  
5   geom_love()
```

Love plot



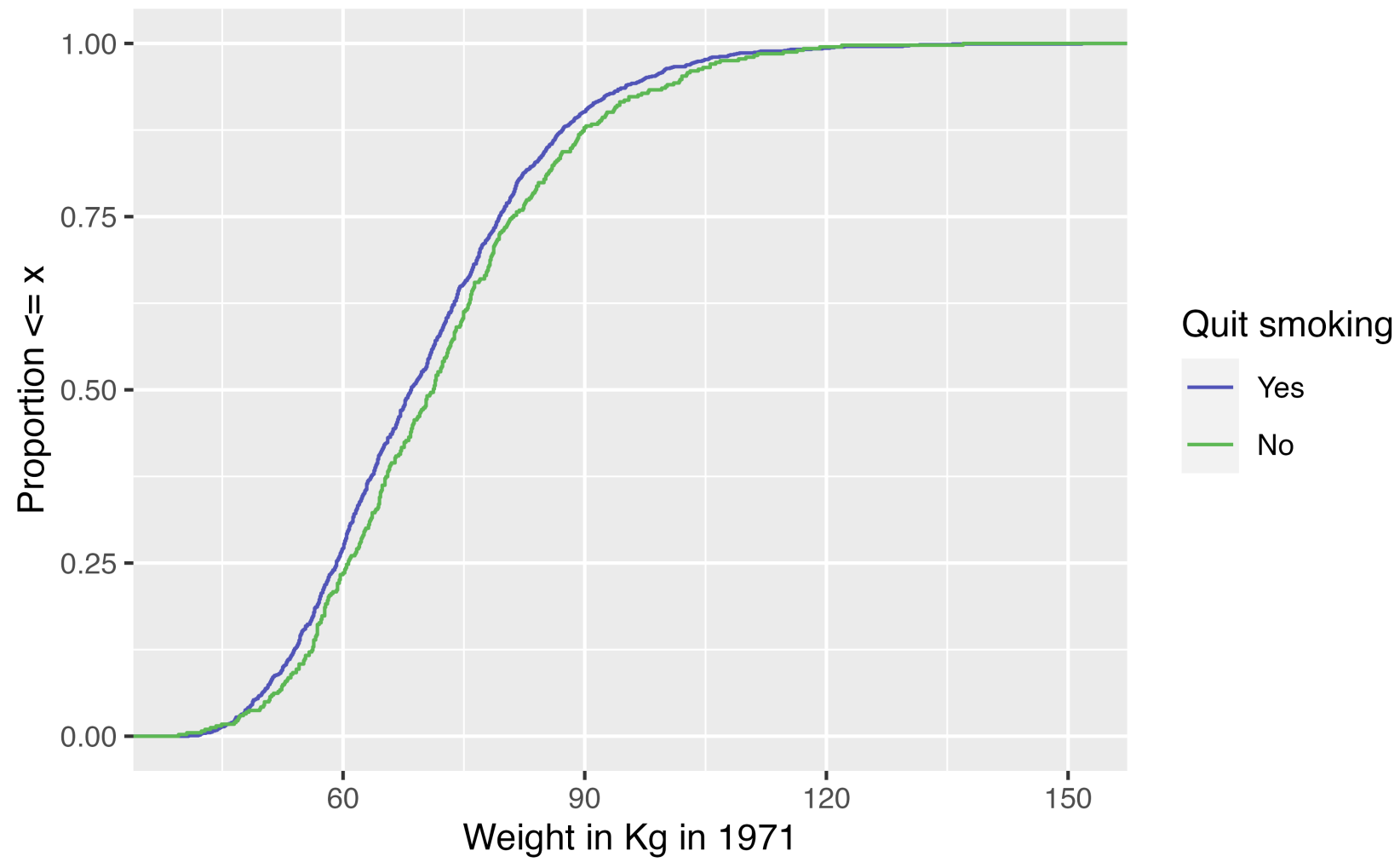
Your turn 1

Create a Love Plot for the propensity score weighting you created in the previous exercise

10:00

ECDF

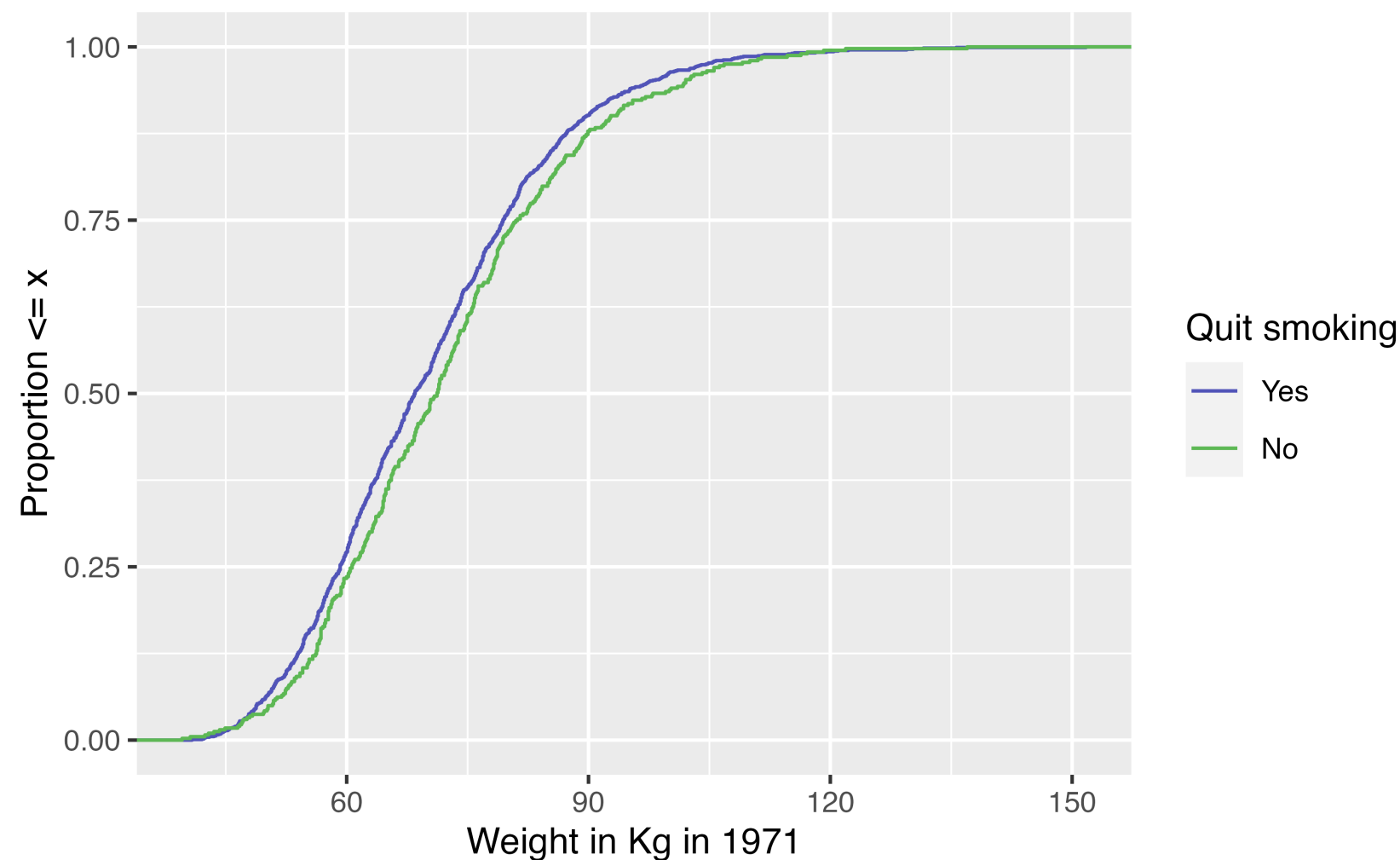
For continuous variables, it can be helpful to look at the *whole* distribution pre and post-weighting rather than a single summary measure



Unweighted ECDF

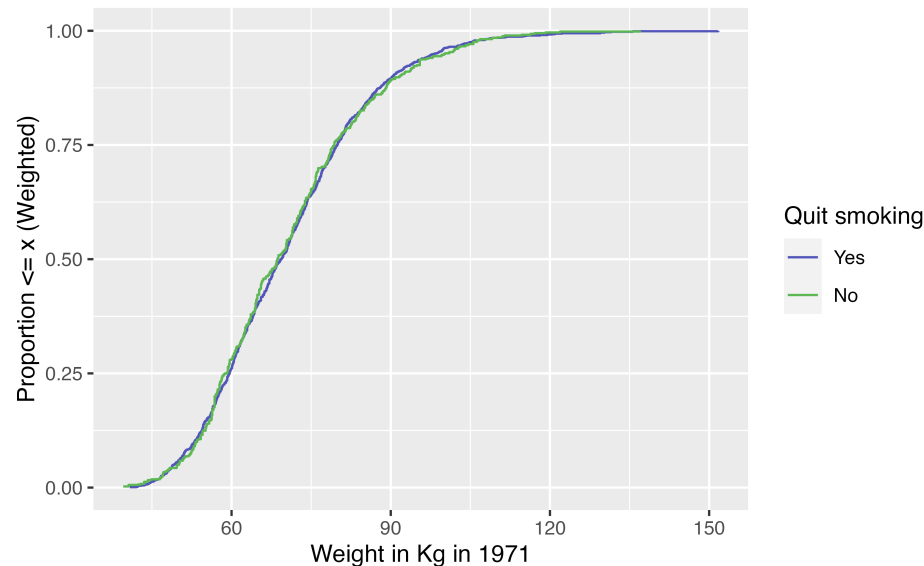
```
1 ggplot(df, aes(x = wt71, color = factor(qsmk))) +  
2   geom_ecdf() +  
3   scale_color_manual(  
4     "Quit smoking",  
5     values = c("#5154B8", "#5DB854"),  
6     labels = c("Yes", "No")  
7   ) +  
8   xlab("Weight in Kg in 1971") +  
9   ylab("Proportion <= x")
```

Unweighted ECDF



Weighted ECDF

```
1 ggplot(df, aes(x = wt71, color = factor(qsmk))) +  
2   geom_ecdf(aes(weights = w_ate)) +  
3   scale_color_manual(  
4     "Quit smoking",  
5     values = c("#5154B8", "#5DB854"),  
6     labels = c("Yes", "No")  
7   ) +  
8   xlab("Weight in Kg in 1971") +  
9   ylab("Proportion <= x (Weighted)")
```



Your turn 2

Create an unweighted ECDF examining the **weather_wdwhigh** confounder by whether or not the day had Extra Magic Hours.

Create a weighted ECDF examining the **weather_wdwhigh** confounder

***Bonus!* Weighted Tables in R**

1. Create a “design object” to incorporate the weights

```
1 library(survey)
2
3 svy_des <- svydesign(
4   ids = ~ 1,
5   data = df,
6   weights = ~ wts
7 )
```

2. Pass to

gtsummary::tbl_svysummary()

```
1 library(gtsummary)
2 tbl_svysummary(svy_des, by = x) |>
3   add_difference(everything() ~ "smd")
4 # modify_column_hide(ci) to hide CI column
```


Characteristic	0, N = 1,565 ¹	1, N = 1,561 ¹	Difference ²
WEIGHT IN KILOGRAMS IN 1971	69 (60, 80)	69 (59, 79)	0.01
0: WHITE 1: BLACK OR OTHER IN 1971			0.01
0	1,359 (87%)	1,352 (87%)	
1	206 (13%)	209 (13%)	
AGE IN 1971	43 (33, 52)	43 (33, 53)	-0.01
0: MALE 1: FEMALE			0.00
0	764 (49%)	764 (49%)	
1	802 (51%)	797 (51%)	
NUMBER OF CIGARETTES SMOKED PER DAY IN 1971	20 (10, 25)	20 (10, 30)	0.02
YEARS OF SMOKING	24 (15, 33)	24 (14, 33)	0.00
IN RECREATION, HOW MUCH EXERCISE? IN 1971, 0:much exercise,1:moderate exercise,2:little or no exercise			0.04
0	302 (19%)	294 (19%)	
1	665 (42%)	691 (44%)	
2	599 (38%)	576 (37%)	
IN YOUR USUAL DAY, HOW ACTIVE ARE YOU? IN 1971, 0:very active, 1:moderately active, 2:inactive			0.03
0	700 (45%)	684 (44%)	
1	718 (46%)	738 (47%)	
2	147 (9.4%)	138 (8.9%)	

¹ Median (IQR); n (%)

² Standardized Mean Difference

