Propensity Score Diagnostics

Lucy D'Agostino McGowan

Wake Forest University

2021-09-01 (updated: 2022-07-23)

Checking balance

Love plots (Standardized Mean Difference)

ECDF plots

Standardized Mean Difference (SMD)

$$d = \frac{\bar{x}_{treatment} - \bar{x}_{control}}{\sqrt{\frac{s^2_{treatment}+s^2_{control}}{2}}}$$

SMD in R

1 Calculate standardized mean differences

```
library(tidysmd)
library(tidyverse)

smds <- tidy_smd(
  df,
  .vars = c(confounder_1, confounder_2, ...),
  .group = exposure,
  .wts = wts # weight is optional
)
```

SMD in R

1 Calculate standardized mean differences

```r
library(tidysmd)
library(tidyverse)

smds <- tidy_smd(
  df,
  .vars = c(confounder_1, confounder_2, ...),
  .group = exposure,
  .wts = wts # weight is optional
)
```

SMD in R

1 Calculate standardized mean differences

```
library(tidysmd)
library(tidyverse)

smds <- tidy_smd(
  df,
  .vars = c(confounder_1, confounder_2, ...),
  .group = exposure,
  .wts = wts # weight is optional
)
```

SMD in R

1 Calculate standardized mean differences

```
library(tidysmd)
library(tidyverse)

smds <- tidy_smd(
  df,
  .vars = c(confounder_1, confounder_2, ...),
  .group = exposure,
  .wts = wts # weight is optional
)
```

SMD in R

1 Calculate standardized mean differences

```
library(tidysmd)
library(tidyverse)

smds <- tidy_smd(
  df,
  .vars = c(confounder_1, confounder_2, ...),
  .group = exposure,
  .wts = wts # weight is optional
)
```

SMD in R

2 Plot them! (in a Love plot!)

```
ggplot(
  data = smds,
  aes(x = abs(smd), y = variable, group = weights, color = weights)
) +
  geom_line(orientation = "y") +
  geom_point() +
  geom_vline(xintercept = 0.1, color = "black", size = 0.1)
```

SMD in R

2 Plot them! (in a Love plot!)

```
ggplot(
  data = smds,
  aes(x = abs(smd), y = variable, group = weights, color = weights)
) +
  geom_line(orientation = "y") +
  geom_point() +
  geom_vline(xintercept = 0.1, color = "black", size = 0.1)
```

SMD in R

2 Plot them! (in a Love plot!)

```r
ggplot(
  data = smds,
  aes(x = abs(smd), y = variable, group = weights, color = weights)
) +
  geom_line(orientation = "y") +
  geom_point() +
  geom_vline(xintercept = 0.1, color = "black", size = 0.1)
```

SMD in R

2 Plot them! (in a Love plot!)

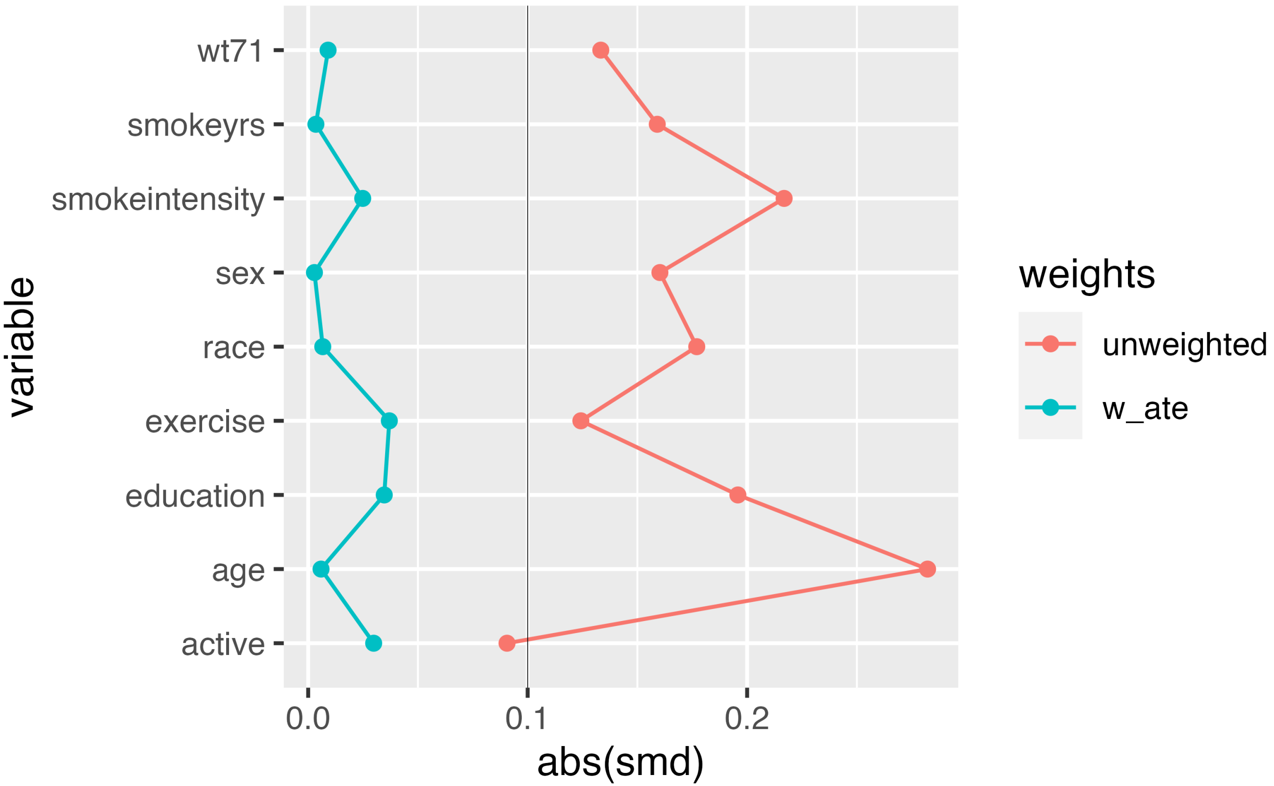
```
ggplot(
  data = smds,
  aes(x = abs(smd), y = variable, group = weights, color = weights)
) +
  geom_line(orientation = "y") +
  geom_point() +
  geom_vline(xintercept = 0.1, color = "black", size = 0.1)
```

SMD in R

2 Plot them! (in a Love plot!)

```
ggplot(
  data = smds,
  aes(x = abs(smd), y = variable, group = weights, color = weights)
) +
  geom_line(orientation = "y") +
  geom_point() +
  geom_vline(xintercept = 0.1, color = "black", size = 0.1)
```

Love plot



wt71
smokeyrs
smokeintensity
sex
race
exercise
education
age
active
variable
0.0
0.1
0.2
abs(smd)
weights
unweighted
w_ate

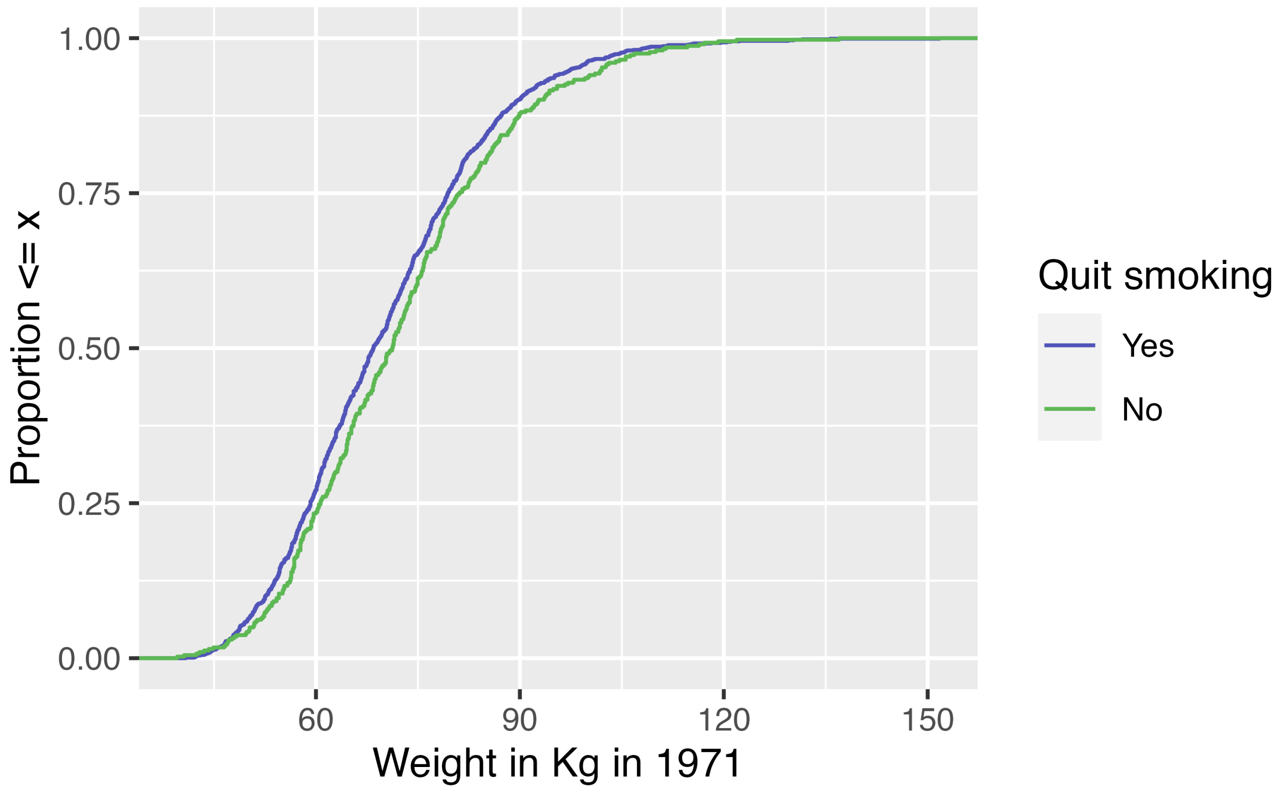
Your turn 1

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

```r
library(ggecdf)

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

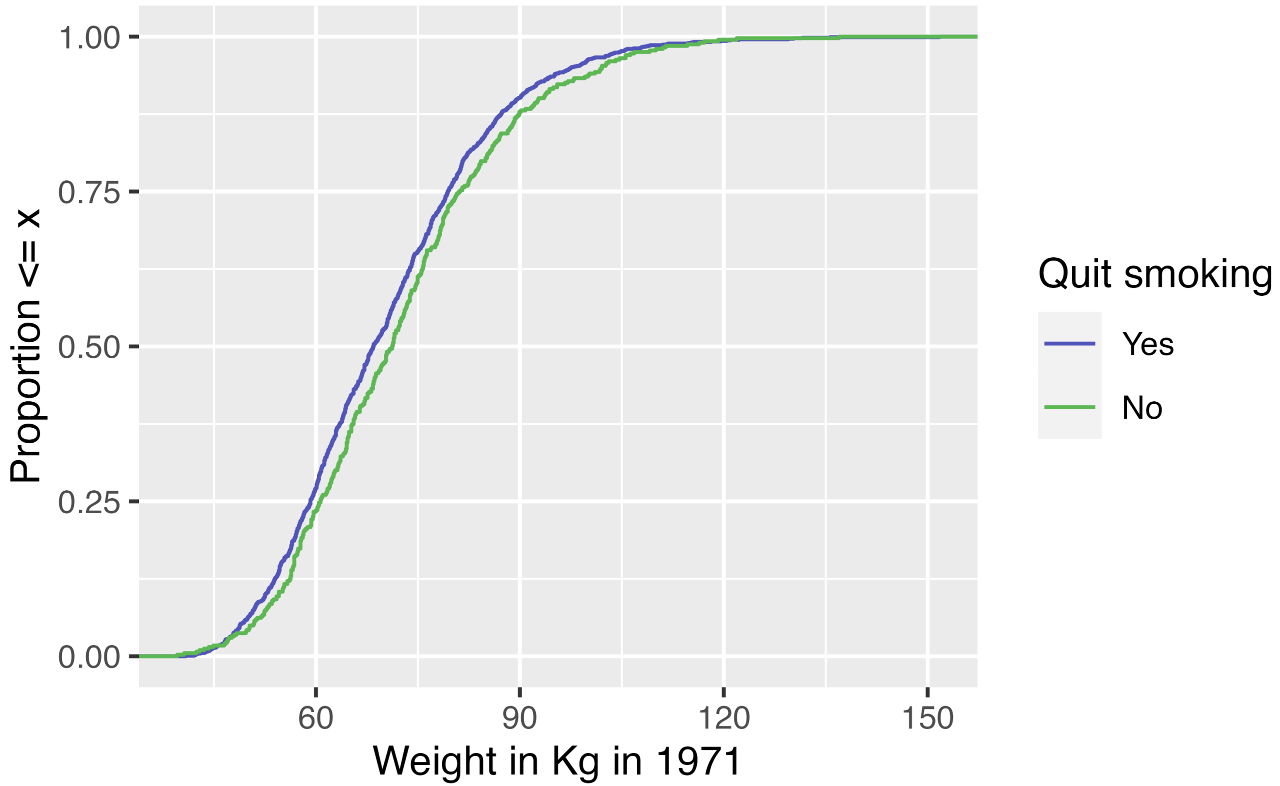
Unweighted ECDF

```r
library(ggecdf)

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

Unweighted ECDF

```
library(ggecdf)

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

Unweighted ECDF



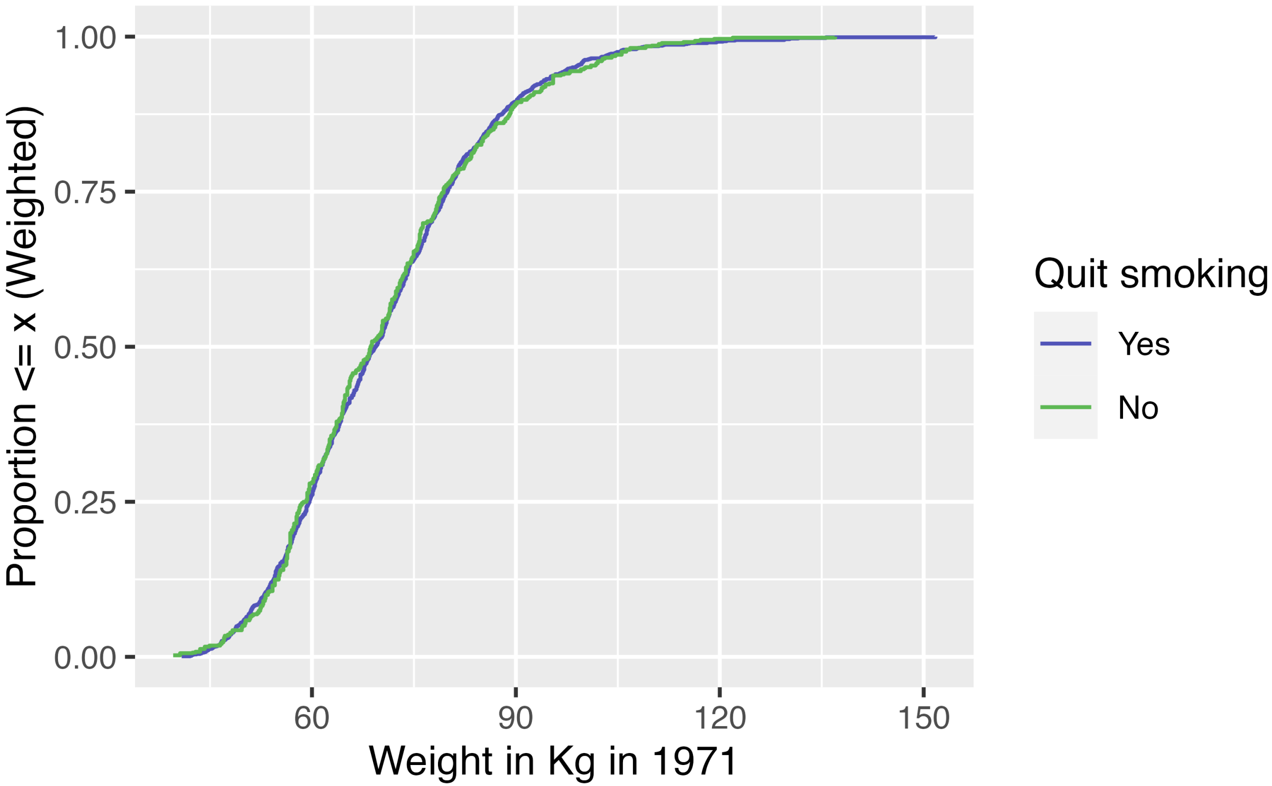
Weighted ECDF

```r
library(ggecdf)

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

Weighted ECDF

```r
library(ggecdf)

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

Weighted ECDF



Your turn 2

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

10:00

Bonus! Weighted Tables in R

Weighted Tables in R

1 Create a "design object" to incorporate the weights

```
library(survey)

svy_des <- svydesign(
  ids = ~ 1,
  data = df,
  weights = ~ wts
)
```

Weighted Tables in R

2 **Pass to** gtsummary::tbl_svysummary()

```
library(gtsummary)
tbl_svysummary(svy_des, by = x) %>%
  add_difference(everything() ~ "smd")
# 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