Regression Discontinuity Simulation Homework

Jennifer Hill, Ray Lu & Zarni Htet

Objective

The goal of this exercise is to simulate and analyze data that might have arisen from a policy where eligibility was determined based on one observed measure. Data from this type of setting are often consistent with the assumptions of the regression discontinuity designs we discussed in class.

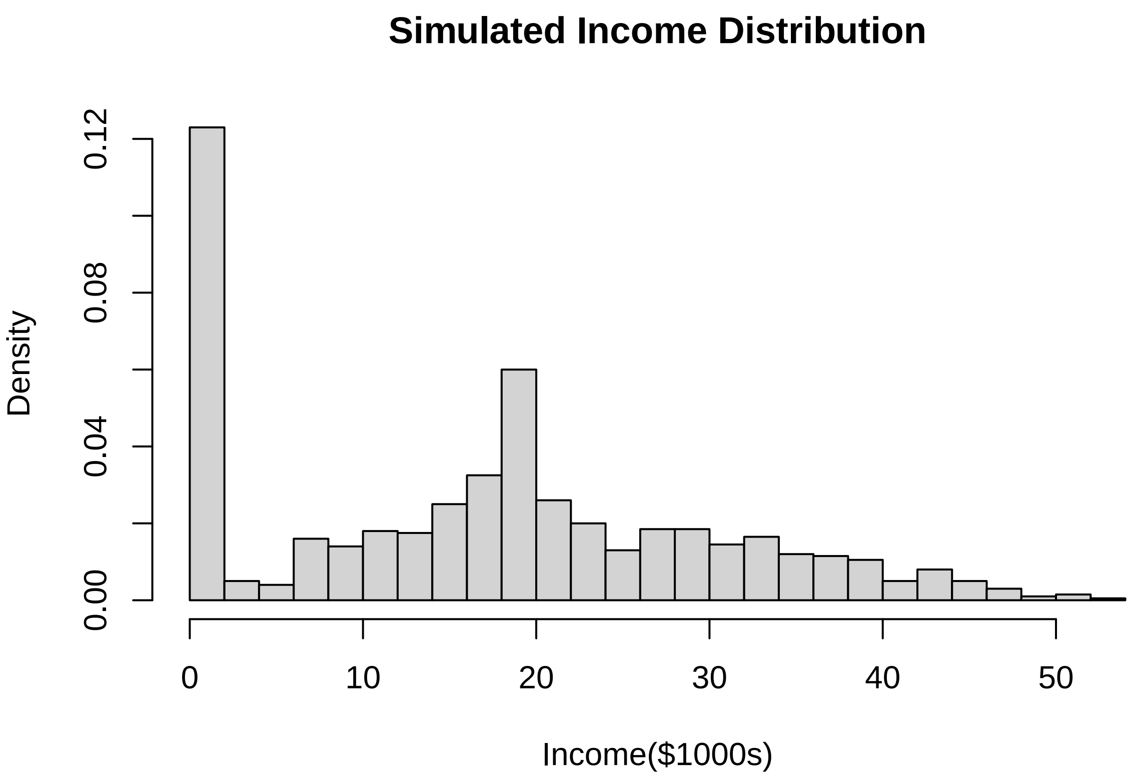
Setting

This assignment simulates hypothetical data collected on women who gave birth at any one of several hospitals in disadvantaged neighborhoods in New York City in 2010. We are envisioning a government policy that makes available pre- and post-natal (through 2 years post-birth) health care for pregnant women, new mothers and their children. This program is only available for women in households with income below $20,000 at the time they gave birth. The general question of interest is whether this program increases a measure of child health at age 3. You will generate data for a sample of 1000 individuals.

Clean regression discontinuity design. For this assignment we will make the unrealistic assumption that everyone who is eligible for the program participates and no one participates who is not eligible.

Question 1. God role: simulate income. Simulate the "running variable" (sometimes referred to as the "running variable", "assignment variable", or "rating"), income, in units of thousands of dollars. Call the variable "income". Try to create a distribution that mimics the key features of the data displayed in income_hist.pdf. Plot your income variable in a histogram with 30 bins.

```r
# Setting seed
set.seed(0)
# Using uniform distributions for the two large peaks, then using normal distribution for the perceived
income <- c(runif(220,0,1),
            runif(70,19,20),
            rnorm(145,12,10),
            rnorm(145,16,5),
            rnorm(270,23,9),
            rnorm(150,35,8)
            )
# Any values less than 0 were set to 0
income <- ifelse(income<0,0,income)
# Histogram of simulated data
hist(income,breaks=30,freq=FALSE,xlab="Income($1000s)",main="Simulated Income Distribution")
```



Question 2. Policy maker role: Assign eligibility indicator. Create an indicator for program eligibility for this sample. Call this variable "eligible".

```
eligible <- as.factor(ifelse(income<20,1,0))
```

Question 3: God role. For question 3 you will simulate a health measure what will serve as the outcome. You will simulate data from *two* possible worlds that vary with regard to the relationships between health (outcome) and income (running variable).

Question 3a

(a) God role. Simulate potential outcomes for World A.

i) Generate the potential outcomes for health assuming linear models for both E[Y(0) | X] and E[Y(1) | X]. This health measure should have a minimum possible score of 0 and maximum possible score of 30. The *expected* treatment effect for everyone should be 4 (in other words, E[Y(1) - Y(0) | X] should be 4 at all levels of X). The residual standard deviation of each potential outcome should be *1*.

ii) Save two datasets: (1) fullA should have the running variable and both potential outcomes and (2) obsA should have the running variable, the eligibility variable, and the observed outcome.

```
# Setting seed for the chunk
set.seed(0)
# Initialize the data frame
fullA <- data.frame(income=income,y0=NA,y1=NA)
# y0 depends on income and is linear, y0 can't be greater than 26 and can't be lower than 0. With sd=1,
# max(income) = 63.4, so to be safe we'll assume the biggest simulated value is 100. 25/100 = 0.25. Aft
fullA$y0 <- 5 + 0.25*income + rnorm(1000,0,1)
summary(fullA$y0)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
```

```
##   2.906   6.383   9.219   9.243  11.558  18.593
```

```r
fullA$y1 <- fullA$y0 + rnorm(1000,4,1)
summary(fullA$y1)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   5.577  10.419  13.126  13.219  15.814  23.882
```

```r
# Check difference in means
mean(fullA$y1-fullA$y0)
```

```
## [1] 3.975214
```

```r
# Create obsA
obsA <- data.frame(income=income,eligible=eligible,y=ifelse(eligible==1,fullA$y1,fullA$y0))
```

Question 3b

(b) Simulate potential outcomes for World B.

i) Generate the potential outcomes for health assuming a linear model for E[Y(0) | X] and a quadratic model for E[Y(1) | X]. The treatment effect at the threshold (the level of X that determines eligibility) should be 4. The residual standard deviation of each potential outcome should be *1.* This health measure should have a minimum possible score of 0 and maximum possible score of 100. Creating this DGP may be facilitated by using a transformed version of your income variable that subtracts out the threshold value.

ii) Save two datasets: (1) fullB should have the running variable and both potential outcomes and (2) obsB should have the running variable, the eligibility variable, and the observed outcome.

```r
# Set seed for chunk
set.seed(0)
# Initialize new data frame for world B
fullB <- data.frame(income=income,y0=NA,y1=NA)
# Transform income subtracting the threshold value, 20
t.income <- income - 20

# Generate new y0s in this world, should be linear. Using mostly the same model as the world A, but usi
fullB$y0 <- 15 + 0.25*t.income + rnorm(1000,0,1)
summary(fullB$y0)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   7.906  11.383  14.219  14.243  16.558  23.593
```

```r
# Generating y1s, logic for t.income means that if income was below 20, (now it's anything less than 0
fullB$y1 <- ifelse(t.income < 0,fullB$y0 + rnorm(1000,4,1),0.07*t.income^2 + rnorm(1000,15,1))
summary(fullB$y1)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.58   15.02   17.03   20.23   19.98   87.10
```

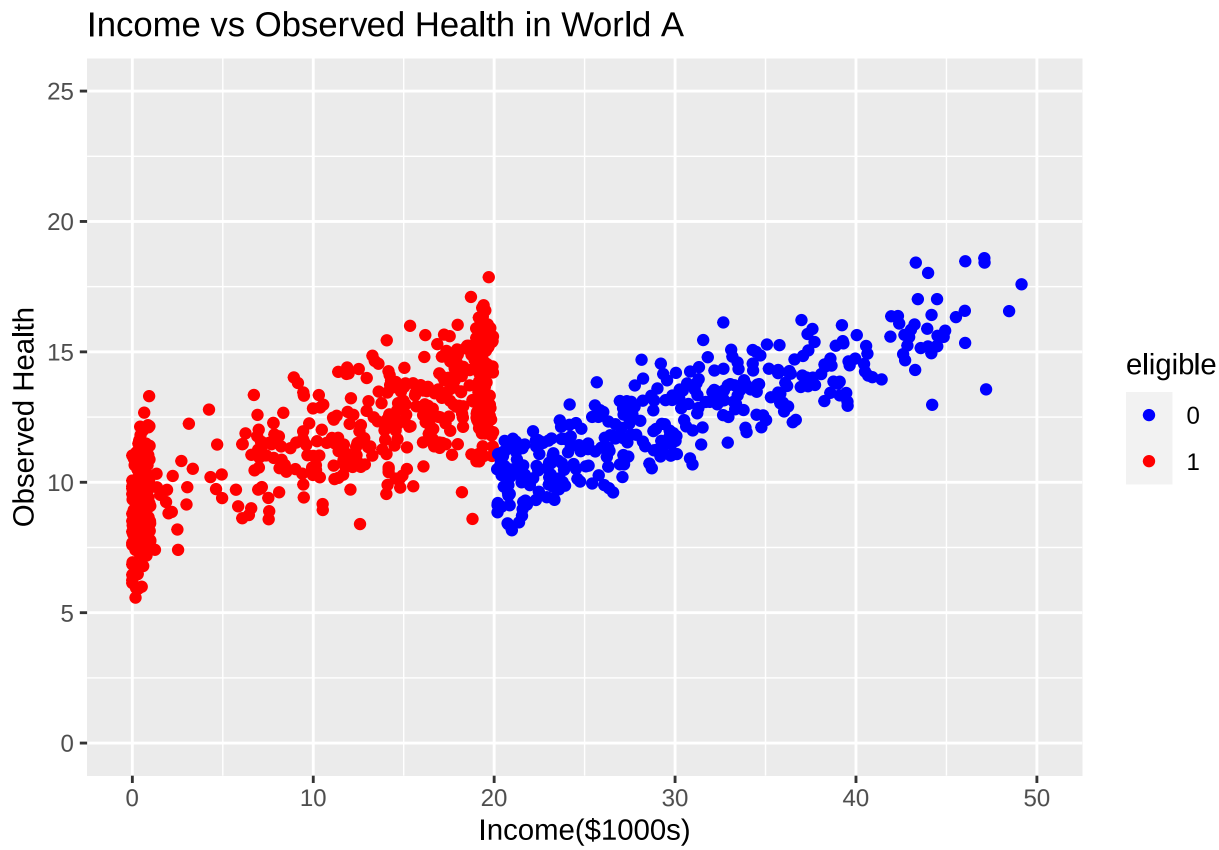
```r
# Create obsB
obsB <- data.frame(income=income,eligible=eligible,y=ifelse(eligible==1,fullB$y1,fullB$y0))
```

Question 4. Researcher role. Plot your data! Make two scatter plots of income (x-axis) versus observed health (y-axis), one corresponding to each world. In each, plot eligible participants in red and non-eligible participants in blue.

```r
# Call ggplot2
library(ggplot2)
```

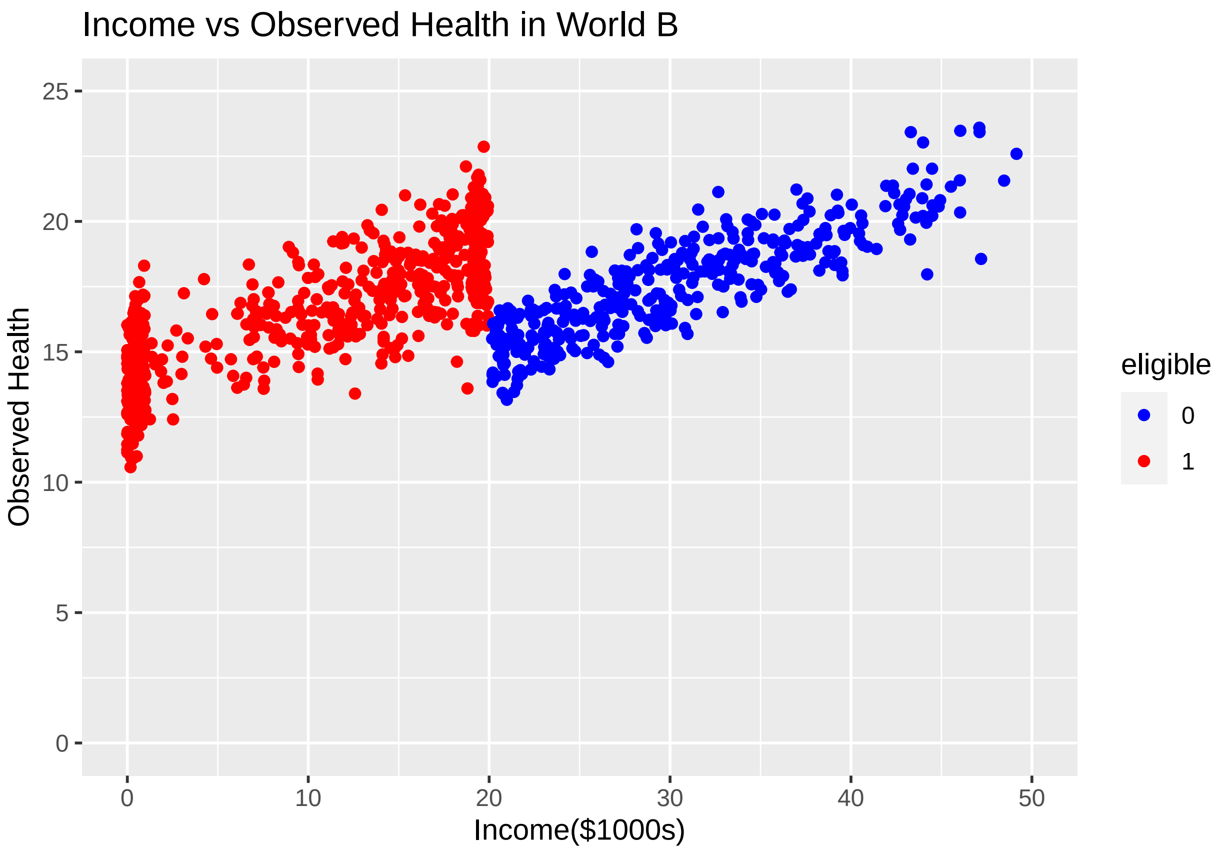
```r
# Plot world A
ggplot(data=obsA) +
  ggtitle("Income vs Observed Health in World A") +
  xlab("Income($1000s)") +
  ylab("Observed Health") +
  xlim(0,50) +
  ylim(0,25) +
  geom_point(aes(x=income,y=y,color=eligible)) +
  scale_color_manual(values=c("blue","red"))
```

```
## Warning: Removed 4 rows containing missing values (geom_point).
```




```r
# Plot world B
ggplot(data=obsB) +
  ggtitle("Income vs Observed Health in World B") +
  xlab("Income($1000s)") +
  ylab("Observed Health") +
  xlim(0,50) +
  ylim(0,25) +
  geom_point(aes(x=income,y=y,color=eligible)) +
  scale_color_manual(values=c("blue","red"))
```

```
## Warning: Removed 4 rows containing missing values (geom_point).
```



Question 5. Researcher role. Estimate the treatment effect for World A and World B using all the data. Now we will estimate effects in a number of different ways. Each model should include reported income and eligible as predictors. In each case use the model fit to report the estimate of the effect of the program at the threshold level of income. All models in Question 5 will be fit to all the data.

Question 5a: Researcher role. Estimates for World A using all the data.

(a) *Using all the data from World A,* perform the following analyses.
(b) Fit a linear model. Do not include an interaction.

(ii) Fit a linear model and include an interaction between income and eligible.
(iii) Fit a model that is quadratic in income and includes an interaction between both income terms and eligible (that is – allow the shape of the relationship to vary between treatment and control groups.

```
# For each model, fit estimate with predict, then calculate the treatment effect. Note the extra argume
fitA1 <- lm(y ~ income + eligible,obsA)
round(summary(fitA1)$coefficients,4)
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.0745     0.1744 29.0978        0
## income        0.2467     0.0053 46.9378        0
## eligible1     3.9120     0.1412 27.7020        0
```

```
predA1 <- predict(fitA1,data.frame(income=c(20,20),eligible=as.factor(c(0,1))))
estA1 <- predA1[2]-predA1[1]
estA1
```

```
##        2
## 3.911978
```

```r
fitA2 <- lm(y ~ eligible * income,obsA)
round(summary(fitA2)$coefficients,4)
```

```
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)       5.2120     0.2776 18.7766   0.0000
## eligible1         3.7520     0.2883 13.0135   0.0000
## income            0.2422     0.0088 27.5844   0.0000
## eligible1:income  0.0070     0.0110  0.6365   0.5246
```

```r
predA2 <- predict(fitA2,data.frame(income=c(20,20),eligible=as.factor(c(0,1))))
estA2 <- predA2[2]-predA2[1]
estA2
```

```
##        2
## 3.891575
```

```r
fitA3 <- lm(y ~ eligible * income * I(income^2),obsA)
round(summary(fitA3)$coefficients,4)
```

```
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     0.1102     4.5959  0.0240   0.9809
## eligible1                       8.7851     4.5969  1.9111   0.0563
## income                          0.6970     0.4334  1.6081   0.1081
## I(income^2)                    -0.0128     0.0132 -0.9756   0.3295
## eligible1:income               -0.3163     0.4418 -0.7158   0.4743
## eligible1:I(income^2)          -0.0060     0.0172 -0.3462   0.7292
## income:I(income^2)              0.0001     0.0001  0.8947   0.3711
## eligible1:income:I(income^2)    0.0005     0.0004  1.3696   0.1711
```

```r
predA3 <- predict(fitA3,data.frame(income=c(20,20),eligible=as.factor(c(0,1))))
estA3 <- predA3[2]-predA3[1]
estA3
```

```
##        2
## 4.304748
```

Question 5b: Researcher role. Estimates for World B using all the data.

(b) *Using all the data from World B,* perform the following analyses.
(c) Fit a linear model. Do not include an interaction.

(ii) Fit a linear model and include an interaction between income and eligible.
(iii) Fit a model that is quadratic in income and includes an interaction between both income terms and eligible (that is – allow the shape of the relationship to vary between the treatment and control groups).

```r
# For each model, fit estimate with predict, then calculate the treatment effect. Note the extra argume
fitB1 <- lm(y ~ income + eligible,obsB)
round(summary(fitB1)$coefficients,4)
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  10.0745     0.1744 57.7682        0
## income        0.2467     0.0053 46.9378        0
## eligible1     3.9120     0.1412 27.7020        0
```

```r
predB1 <- predict(fitB1,data.frame(income=c(20,20),eligible=as.factor(c(0,1))))
estB1 <- predB1[2]-predB1[1]
estB1
```

```
##        2
```

```
## 3.911978
```

```
fitB2 <- lm(y ~ eligible * income,obsB)
round(summary(fitB2)$coefficients,4)
```

```
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      10.2120     0.2776 36.7896   0.0000
## eligible1         3.7520     0.2883 13.0135   0.0000
## income            0.2422     0.0088 27.5844   0.0000
## eligible1:income  0.0070     0.0110  0.6365   0.5246
```

```
predB2 <- predict(fitB2,data.frame(income=c(20,20),eligible=as.factor(c(0,1))))
estB2 <- predB2[2]-predB2[1]
estB2
```

```
##        2
## 3.891575
```

```
fitB3 <- lm(y ~ eligible * income * I(income^2),obsB)
round(summary(fitB3)$coefficients,4)
```

```
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    5.1102     4.5959  1.1119   0.2664
## eligible1                      8.7851     4.5969  1.9111   0.0563
## income                         0.6970     0.4334  1.6081   0.1081
## I(income^2)                   -0.0128     0.0132 -0.9756   0.3295
## eligible1:income              -0.3163     0.4418 -0.7158   0.4743
## eligible1:I(income^2)         -0.0060     0.0172 -0.3462   0.7292
## income:I(income^2)             0.0001     0.0001  0.8947   0.3711
## eligible1:income:I(income^2)   0.0005     0.0004  1.3696   0.1711
```

```
predB3 <- predict(fitB3,data.frame(income=c(20,20),eligible=as.factor(c(0,1))))
estB3 <- predB3[2]-predB3[1]
estB3
```

```
##        2
## 4.304748
```

Question 6. Researcher role. Estimate the treatment effect for World A and World B using data close to the threshold. We will again estimate effects in a number of different ways. Each model should include "income" and "eligible" as predictors. In each case use the model fit to report the estimate of the effect of the program at the threshold level of income. All models in Question 6 will be fit only to women with incomes ranging from \$18,000 to \$22,000.

Question 6a: Researcher role. Estimates for World A using the restricted data.

(a) Using the restricted data (for participants with incomes between \$18K and \$22K) from World A, perform the following analyses.
(b) Fit a linear model to the restricted dataset. Do not include an interaction.

(ii) Fit a linear model to the restricted dataset, include an interaction between income and eligible.
(iii) Fit a model that is quadratic in income and includes an interaction between both income terms and eligible (that is – allow the shape of the relationship to vary between the treatment and control groups).

```
# For each model, fit estimate with predict, then calculate the treatment effect. Note the extra argume
restrict <- (income <= 22 & income >=18)
fitA1r <- lm(y ~ income + eligible,obsA,subset=restrict)
round(summary(fitA1r)$coefficients,4)
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   6.3330     4.8763  1.2987   0.1958
## income        0.1818     0.2321  0.7833   0.4346
## eligible1     3.9455     0.4772  8.2685   0.0000
```

```r
predA1r <- predict(fitA1r,data.frame(income=c(20,20),eligible=as.factor(c(0,1))))
estA1r <- predA1r[2]-predA1r[1]
estA1r
```

```
##        2
## 3.945517
```

```r
fitA2r <- lm(y ~ eligible * income,obsA,subset=restrict)
round(summary(fitA2r)$coefficients,4)
```

```
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       12.1463     8.1556  1.4893   0.1383
## eligible1         -4.8299     9.8765 -0.4890   0.6255
## income            -0.0951     0.3884 -0.2450   0.8068
## eligible1:income   0.4310     0.4845  0.8896   0.3750
```

```r
predA2r <- predict(fitA2r,data.frame(income=c(20,20),eligible=as.factor(c(0,1))))
estA2r <- predA2r[2]-predA2r[1]
estA2r
```

```
##        2
## 3.790194
```

```r
fitA3r <- lm(y ~ eligible * income * I(income^2),obsA,subset=restrict)
round(summary(fitA3r)$coefficients,4)
```

```
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   2100.2730 16835.4978  0.1248   0.9009
## eligible1                    -1819.0015 18824.5365 -0.0966   0.9231
## income                        -310.8528  2405.9392 -0.1292   0.8974
## I(income^2)                     15.3854   114.5752  0.1343   0.8933
## eligible1:income               258.0030  2747.2696  0.0939   0.9253
## eligible1:I(income^2)          -12.0680   134.0505 -0.0900   0.9284
## income:I(income^2)              -0.2534     1.8182 -0.1394   0.8893
## eligible1:income:I(income^2)     0.1862     2.1876  0.0851   0.9323
```

```r
predA3r <- predict(fitA3r,data.frame(income=c(20,20),eligible=as.factor(c(0,1))))
estA3r <- predA3r[2]-predA3r[1]
estA3r
```

```
##        2
## 3.736644
```

Question 6b: Researcher role. Estimates for World B using the restricted data.

(b) Using the restricted data (for participants with incomes between \$18K and \$22K) from World B, perform the following analyses.

(c) Fit a linear model to the restricted dataset. Do not include an interaction.

(ii) Fit a linear model to the restricted dataset, include an interaction between income and eligible.

(iii) Fit a model that is quadratic in income and includes an interaction between both income terms and eligible (that is – allow the shape of the relationship to vary between treatment and control groups.

```
# For each model, fit estimate with predict, then calculate the treatment effect. Note the extra argume
fitB1r <- lm(y ~ income + eligible,obsB,subset=restrict)
round(summary(fitA1r)$coefficients,4)
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   6.3330     4.8763  1.2987   0.1958
## income        0.1818     0.2321  0.7833   0.4346
## eligible1     3.9455     0.4772  8.2685   0.0000
```

```
predB1r <- predict(fitB1r,data.frame(income=c(20,20),eligible=as.factor(c(0,1))))
estB1r <- predB1r[2]-predB1r[1]
estB1r
```

```
##        2
## 3.945517
```

```
fitB2r <- lm(y ~ eligible * income,obsB,subset=restrict)
round(summary(fitB2r)$coefficients,4)
```

```
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       17.1463     8.1556  2.1024   0.0370
## eligible1         -4.8299     9.8765 -0.4890   0.6255
## income            -0.0951     0.3884 -0.2450   0.8068
## eligible1:income   0.4310     0.4845  0.8896   0.3750
```

```
predB2r <- predict(fitB2r,data.frame(income=c(20,20),eligible=as.factor(c(0,1))))
estB2r <- predB2r[2]-predB2r[1]
estB2r
```

```
##        2
## 3.790194
```

```
fitB3r <- lm(y ~ eligible * income * I(income^2),obsB,subset=restrict)
round(summary(fitB3r)$coefficients,4)
```

```
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   2105.2730 16835.4978  0.1250   0.9006
## eligible1                    -1819.0015 18824.5365 -0.0966   0.9231
## income                        -310.8528  2405.9392 -0.1292   0.8974
## I(income^2)                     15.3854   114.5752  0.1343   0.8933
## eligible1:income               258.0030  2747.2696  0.0939   0.9253
## eligible1:I(income^2)          -12.0680   134.0505 -0.0900   0.9284
## income:I(income^2)              -0.2534     1.8182 -0.1394   0.8893
## eligible1:income:I(income^2)     0.1862     2.1876  0.0851   0.9323
```

```
predB3r <- predict(fitB3r,data.frame(income=c(20,20),eligible=as.factor(c(0,1))))
estB3r <- predB3r[2]-predB3r[1]
estB3r
```

```
##        2
## 3.736644
```

Question 7. Researcher role. Displaying your estimates. Present your estimates from questions 5 and 6 into one or two tables or figures, clearly noting which world the data are from, which models the estimates are from, and which analysis sample was used.

```
sim.results <-
data.frame(
```

```r
  world=c(rep("A",3),rep("B",3),rep("A",3),rep("B",3)),
  model=c(rep(c("income, eligible","income, eligible, interaction","income, eligible, interaction, incom
  sample=c(rep("full sample",6),rep("18k < income < 22k",6)),
  estimate=c(estA1,estA2,estA3,estB1,estB2,estB3,estA1r,estA2r,estA3r,estB1r,estB2r,estB3r)
  )
sim.results
```

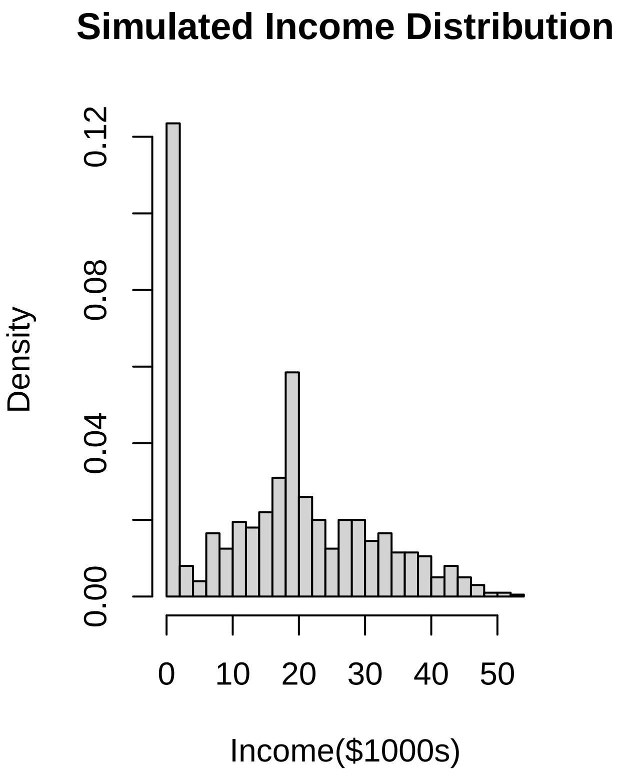
```
##    world                                         model             sample
## 1      A                              income, eligible        full sample
## 2      A                 income, eligible, interaction        full sample
## 3      A income, eligible, interaction, income squared        full sample
## 4      B                              income, eligible        full sample
## 5      B                 income, eligible, interaction        full sample
## 6      B income, eligible, interaction, income squared        full sample
## 7      A                              income, eligible 18k < income < 22k
## 8      A                 income, eligible, interaction 18k < income < 22k
## 9      A income, eligible, interaction, income squared 18k < income < 22k
## 10     B                              income, eligible 18k < income < 22k
## 11     B                 income, eligible, interaction 18k < income < 22k
## 12     B income, eligible, interaction, income squared 18k < income < 22k
##    estimate
## 1  3.911978
## 2  3.891575
## 3  4.304748
## 4  3.911978
## 5  3.891575
## 6  4.304748
## 7  3.945517
## 8  3.790194
## 9  3.736644
## 10 3.945517
## 11 3.790194
## 12 3.736644
```

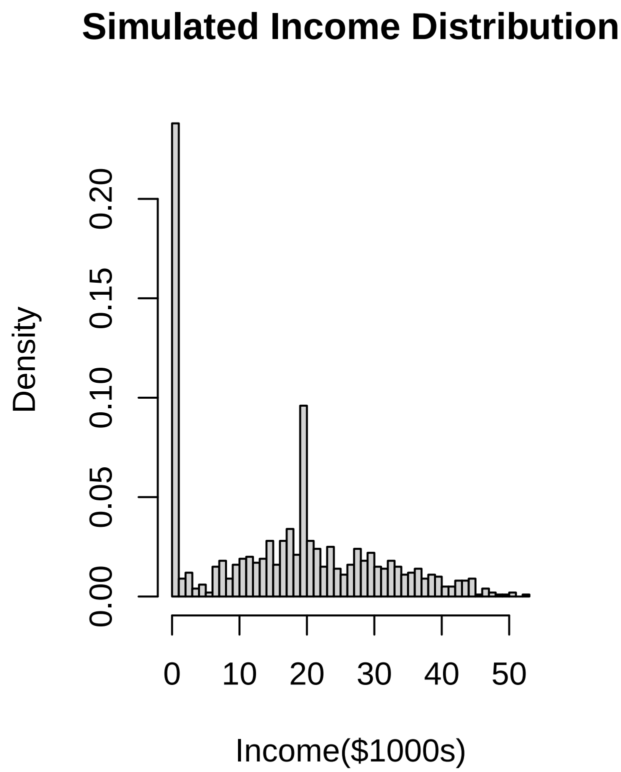
Question 8. Researcher role. Thinking about the data.

(a) A colleague now points out to you that some women may have incentives in these settings to misreport their actual income. Plot a histogram of reported income (using the default settings which should give you 33 bins) and look for anything that might support such a claim. What assumption is called into question if women are truly misreporting in this manner?

```r
# Setting seed
set.seed(0)
# Using uniform distributions for the two large peaks, then using normal distribution for the perceived
income <- c(runif(220,0,1),
            runif(70,20,20),
            rnorm(145,12,10),
            rnorm(145,16,5),
            rnorm(270,23,9),
            rnorm(150,35,8)
            )
# Any values less than 0 were set to 0
income <- ifelse(income<0,0,income)
# Histogram of simulated data
par(mfrow = c(1, 2))
```

```r
hist(income,breaks=30,freq=FALSE,xlab="Income($1000s)",main="Simulated Income Distribution")
hist(income,breaks=60,freq=FALSE,xlab="Income($1000s)",main="Simulated Income Distribution")
```





This calls into question the monotonicity assumption. Women are defying the assignment mechanism by under-reporting their income just enough to qualify for treatment assignment.

(b) Another colleague points out to you that several other government programs (including food stamps and Headstart) have the same income threshold for eligibility. How might this knowledge impact your interpretation of your results?

This impacts the ignoribility assumption, as the same women may qualify for other government programs with the same income threshold, resulting in confounding effects.

Question 9. Researcher role. Thinking about the assumptions? What are the three most important assumptions for the causal estimates in questions 5 and 6?

Ignoribility, monotonicity, and probability of the assignemnt mechanism is related to the probability of receiving treatment.

Question 10. Provide a causal interpretation of your estimate in Question 6biii.

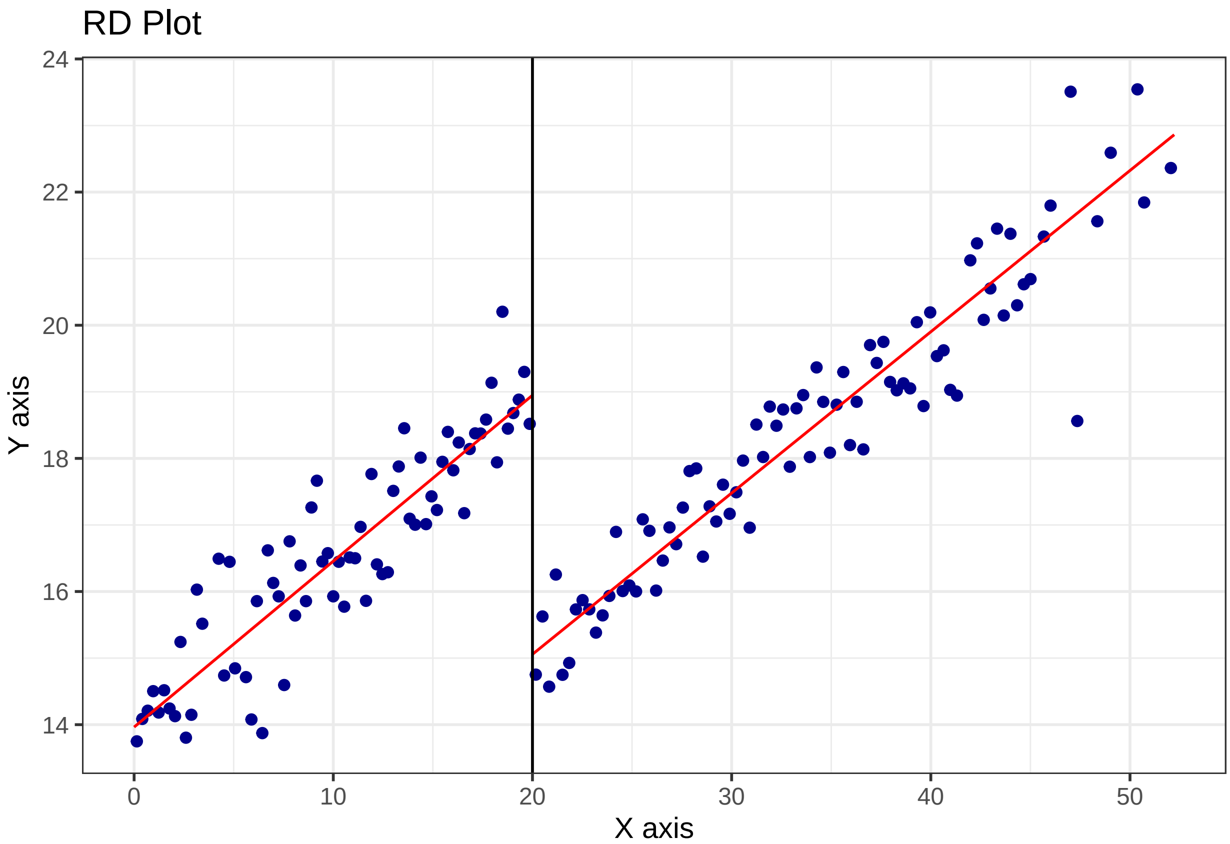
For women who report incomes between 18k and 22k, the treatment effect of program on the health measure is a 3.6 point increase if they were given treatment compared to had they not been given treatment.

Question 11 Use the rdrobust package in R to plot your data and choose an optimal bandwidth for the data in World B.

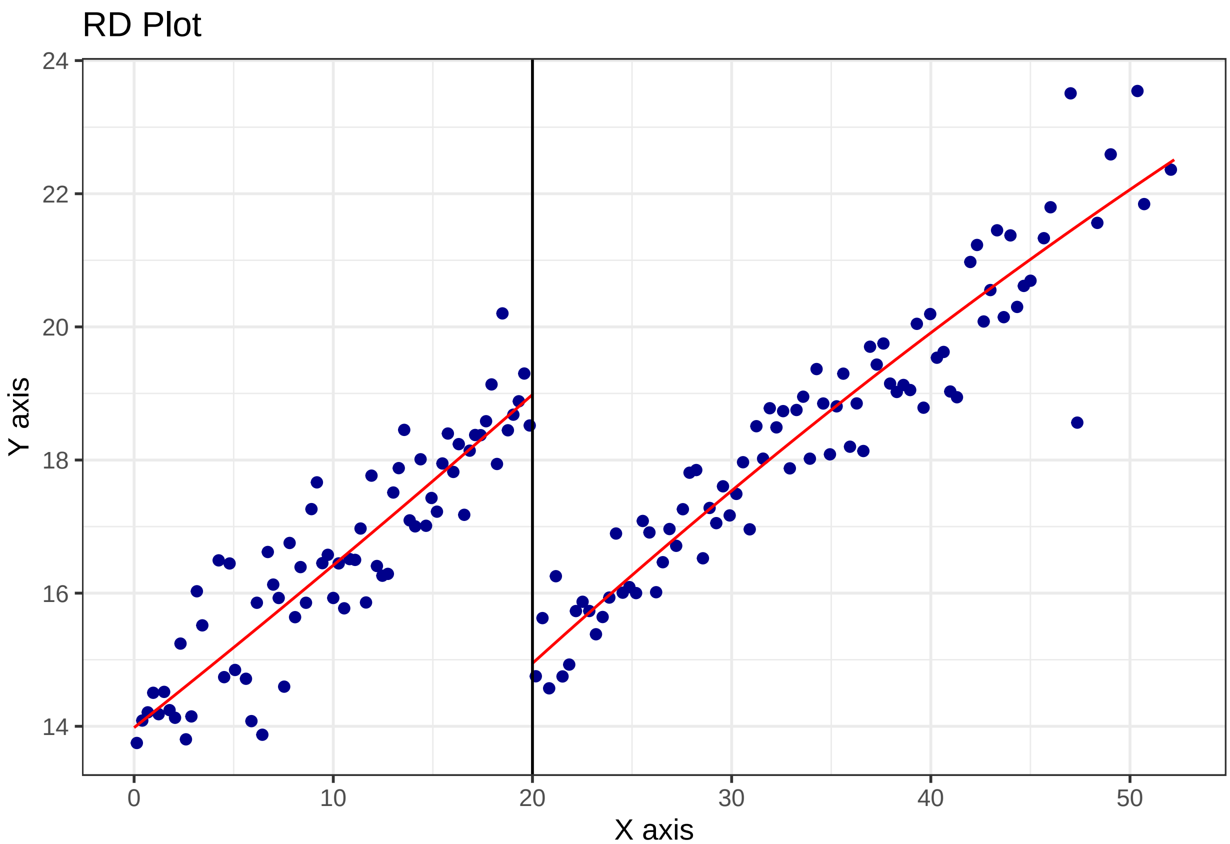
(a) First create two plots using the rdplot command. In the first you use a linear model (that is, use p=1) and in the second allow for a quadratic term (that is, use p=2).

```r
# Loading rdrobust
library(rdrobust)
```

```r
# Specifying cutoff at 20 and p=1 and 2 for linear and quadratic, respectively.
rdplot(y=obsB$y,x=obsB$income,c=20,p=1)
```




```r
rdplot(y=obsB$y,x=obsB$income,c=20,p=2)
```



(b) Now use the rdrobust command to fit an RDD models to your data. Again try both a linear and a quadratic fit. For the bandwidth selection method use "msetwo". [Use the conventional version of estimates. Don't use the "bias corrected"" versions of things.] Compare the points estimates and bandwidths across these approaches. **Extra credit of four points in this question if you also run these commands with a different bandwidth selection method and compare the estimands and the bandwidth endpoints.**

```
# Calling rdrobust model fit to linear model and quadratic model, respectively.
summary(rdrobust(y=obsB$y,x=obsB$income,c=20,p=1,bwselect="msetwo"))
```

```
## Call: rdrobust
##
## Number of Obs.                 1000
## BW type                      msetwo
## Kernel                   Triangular
## VCE method                       NN
##
## Number of Obs.                  630          370
## Eff. Number of Obs.             235          137
## Order est. (p)                    1            1
## Order bias  (q)                   2            2
## BW est. (h)                   6.011        7.056
## BW bias (b)                   8.419       10.740
## rho (h/b)                     0.714        0.657
## Unique Obs.                     617          370
##
## =============================================================================
##         Method     Coef. Std. Err.          z     P>|z|      [ 95% C.I. ]
```

```
## =============================================================================
##    Conventional    -4.071     0.251   -16.195     0.000    [-4.564 , -3.579]
##          Robust         -         -   -13.286     0.000    [-4.612 , -3.426]
## =============================================================================
```

```r
summary(rdrobust(y=obsB$y,x=obsB$income,c=20,p=2,bwselect="msetwo"))
```

```
## Call: rdrobust
##
## Number of Obs.                 1000
## BW type                      msetwo
## Kernel                   Triangular
## VCE method                       NN
##
## Number of Obs.                 630          370
## Eff. Number of Obs.            246          198
## Order est. (p)                   2            2
## Order bias  (q)                  3            3
## BW est. (h)                  6.751       10.314
## BW bias (b)                  9.397       14.241
## rho (h/b)                    0.718        0.724
## Unique Obs.                    617          370
##
## =============================================================================
##          Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## =============================================================================
##    Conventional    -3.946     0.325   -12.158     0.000    [-4.582 , -3.310]
##          Robust         -         -   -10.440     0.000    [-4.604 , -3.148]
## =============================================================================
```

The point estimates for both linear and quadratic are close, but the standard error for the quadratic is greater than the standard error for the linear model, 0.35 versus 0.25. In addition, the bandwidth of the linear is between 6 and 7, whereas the bandwidth for the quadratic is between 6.7 and 10.3. This makes sense given that in a linear fit, fewer points are needed to sufficiently fit a representative model to the data than compared to a quadratic model.

```r
summary(rdrobust(y=obsB$y,x=obsB$income,c=20,p=1,bwselect="mserd"))
```

```
## Call: rdrobust
##
## Number of Obs.                 1000
## BW type                       mserd
## Kernel                   Triangular
## VCE method                       NN
##
## Number of Obs.                 630          370
## Eff. Number of Obs.            231          118
## Order est. (p)                   1            1
## Order bias  (q)                  2            2
## BW est. (h)                  5.927        5.927
## BW bias (b)                  9.014        9.014
## rho (h/b)                    0.658        0.658
## Unique Obs.                    617          370
##
## =============================================================================
##          Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
```

```
## =============================================================================
##   Conventional    -4.052     0.261   -15.555     0.000    [-4.563 , -3.542]
##         Robust         -         -   -12.927     0.000    [-4.585 , -3.378]
## =============================================================================
```

```r
summary(rdrobust(y=obsB$y,x=obsB$income,c=20,p=2,bwselect="mserd"))
```

```
## Call: rdrobust
##
## Number of Obs.                 1000
## BW type                       mserd
## Kernel                   Triangular
## VCE method                       NN
##
## Number of Obs.                 630          370
## Eff. Number of Obs.            248          136
## Order est. (p)                   2            2
## Order bias  (q)                  3            3
## BW est. (h)                  6.953        6.953
## BW bias (b)                  9.995        9.995
## rho (h/b)                    0.696        0.696
## Unique Obs.                    617          370
##
## =============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## =============================================================================
##   Conventional    -3.881     0.345   -11.257     0.000    [-4.557 , -3.206]
##         Robust         -         -    -9.713     0.000    [-4.560 , -3.028]
## =============================================================================
```

```r
summary(rdrobust(y=obsB$y,x=obsB$income,c=20,p=1,bwselect="msesum"))
```

```
## Call: rdrobust
##
## Number of Obs.                 1000
## BW type                      msesum
## Kernel                   Triangular
## VCE method                       NN
##
## Number of Obs.                 630          370
## Eff. Number of Obs.            236          118
## Order est. (p)                   1            1
## Order bias  (q)                  2            2
## BW est. (h)                  6.025        6.025
## BW bias (b)                  8.661        8.661
## rho (h/b)                    0.696        0.696
## Unique Obs.                    617          370
##
## =============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## =============================================================================
##   Conventional    -4.060     0.259   -15.699     0.000    [-4.567 , -3.553]
##         Robust         -         -   -12.776     0.000    [-4.586 , -3.366]
## =============================================================================
```

```r
summary(rdrobust(y=obsB$y,x=obsB$income,c=20,p=2,bwselect="msesum"))
```

```
## Call: rdrobust
##
## Number of Obs.                   1000
## BW type                        msesum
## Kernel                     Triangular
## VCE method                         NN
##
## Number of Obs.                   630          370
## Eff. Number of Obs.              259          148
## Order est. (p)                     2            2
## Order bias  (q)                    3            3
## BW est. (h)                    7.515        7.515
## BW bias (b)                   10.165       10.165
## rho (h/b)                      0.739        0.739
## Unique Obs.                      617          370
##
## =============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## =============================================================================
##   Conventional    -3.936     0.332   -11.844     0.000    [-4.587 , -3.284]
##         Robust         -         -   -10.039     0.000    [-4.581 , -3.084]
## =============================================================================
```

```r
summary(rdrobust(y=obsB$y,x=obsB$income,c=20,p=1,bwselect="msecomb1"))
```

```
## Call: rdrobust
##
## Number of Obs.                   1000
## BW type                      msecomb1
## Kernel                     Triangular
## VCE method                         NN
##
## Number of Obs.                   630          370
## Eff. Number of Obs.              231          118
## Order est. (p)                     1            1
## Order bias  (q)                    2            2
## BW est. (h)                    5.927        5.927
## BW bias (b)                    8.661        8.661
## rho (h/b)                      0.684        0.684
## Unique Obs.                      617          370
##
## =============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## =============================================================================
##   Conventional    -4.052     0.261   -15.555     0.000    [-4.563 , -3.542]
##         Robust         -         -   -12.737     0.000    [-4.583 , -3.360]
## =============================================================================
```

```r
summary(rdrobust(y=obsB$y,x=obsB$income,c=20,p=2,bwselect="msecomb1"))
```

```
## Call: rdrobust
##
## Number of Obs.                   1000
```

```
## BW type                      msecomb1
## Kernel                     Triangular
## VCE method                         NN
##
## Number of Obs.                    630          370
## Eff. Number of Obs.               248          136
## Order est. (p)                      2            2
## Order bias  (q)                     3            3
## BW est. (h)                     6.953        6.953
## BW bias (b)                     9.995        9.995
## rho (h/b)                       0.696        0.696
## Unique Obs.                       617          370
##
## =============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## =============================================================================
##   Conventional    -3.881     0.345   -11.257     0.000    [-4.557 , -3.206]
##         Robust         -         -    -9.713     0.000    [-4.560 , -3.028]
## =============================================================================
```

```
summary(rdrobust(y=obsB$y,x=obsB$income,c=20,p=1,bwselect="cerrd"))
```

```
## Call: rdrobust
##
## Number of Obs.                   1000
## BW type                         cerrd
## Kernel                     Triangular
## VCE method                         NN
##
## Number of Obs.                    630          370
## Eff. Number of Obs.               187           95
## Order est. (p)                      1            1
## Order bias  (q)                     2            2
## BW est. (h)                     4.196        4.196
## BW bias (b)                     9.014        9.014
## rho (h/b)                       0.465        0.465
## Unique Obs.                       617          370
##
## =============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## =============================================================================
##   Conventional    -3.933     0.296   -13.293     0.000    [-4.513 , -3.353]
##         Robust         -         -   -11.919     0.000    [-4.531 , -3.251]
## =============================================================================
```

```
summary(rdrobust(y=obsB$y,x=obsB$income,c=20,p=2,bwselect="cerrd"))
```

```
## Call: rdrobust
##
## Number of Obs.                   1000
## BW type                         cerrd
## Kernel                     Triangular
## VCE method                         NN
##
## Number of Obs.                    630          370
```

```
## Eff. Number of Obs.             195         101
## Order est. (p)                    2           2
## Order bias  (q)                   3           3
## BW est. (h)                   4.685       4.685
## BW bias (b)                   9.995       9.995
## rho (h/b)                     0.469       0.469
## Unique Obs.                     617         370
##
## =============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## =============================================================================
##   Conventional    -3.690     0.447    -8.257     0.000    [-4.566 , -2.814]
##         Robust         -         -    -7.884     0.000    [-4.571 , -2.751]
## =============================================================================
```

The other methods of bandwidth selection strangely end up with single points as bandwidths. Not sure why. I tried.