Election and Conditional Cash Transfer Program in Mexico

October 12, 2022

In this HW, we analyze the data from a study that seeks to estimate the electoral impact of 'Progresa', Mexico's *conditional cash transfer program* (CCT program). This program has been the model for similar programs implemented in many countries around the world where the government provides cash to low income families conditionally on their taking some required actions. For the Progresa program the required actions involved attending workshops regarding health behaviors and having children, particularly girls, attend school. The impacts of the program on the intended outcomes, socioeconomic status and intergenerational transfer of poverty, are strong. Here the interest is in a possible positive side-effect of the program on improving voter turnout and perhaps impacting which party citizens voted for.

This exercise is based on the following two articles:

- Ana de la O. (2013). 'Do Conditional Cash Transfers Affect Voting Behavior? Evidence from a Randomized Experiment in Mexico.' *American Journal of Political Science*, 57:1, pp.1-14; and
- Kosuke Imai, Gary King, and Carlos Velasco. (2015). 'Do Nonpartisan Programmatic Policies Have Partisan Electoral Effects? Evidence from Two Large Scale Randomized Experiments.' Working Paper.

The original study relied on a randomized evaluation of the CCT program in which eligible villages were randomly assigned to receive the program either 21 months (Early *Progresa*) or 6 months (Late *Progresa*) before the 2000 Mexican presidential election. The government did not have the resources to provide the *Progresa* program to all eligible villages when it was first decided upon. For this reason, the government decided it was ethical to randomize provision of the program to begin early or later - this is an example of a lagged randomized study design. The treatment is Early *Progresa* and the alternative is Late *Progresa*.

The author of the original study hypothesized that the CCT program would mobilize voters, leading to an increase in turnout and more support for the incumbent party (PRI in this case). The analysis was based on a sample of precincts that each contain at most one village participating in the evaluation.

The data we analyze are available as the CSV file `progresa.csv`. The names and descriptions of variables in the data set are:

Name	Description
`treatment`	Whether an electoral precinct contains a village where households received Early *Progresa*
`pri2000s`	PRI votes in the 2000 election as a share of precinct population above 18 (in percentage points)
`t2000`	Turnout in the 2000 election as a share of precinct population above 18 (in percentage points)
`t1994`	Turnout in the 1994 election as a share of precinct population above 18 (in percentage points)
`avgpoverty`	Precinct Avg of Village Poverty Index
`pobtot1994`	Total Population in the precinct
`villages`	Number of villages in the precinct

Each observation in the data represents a precinct, and for each precinct the file contains information about its treatment status, the outcomes of interest, socioeconomic indicators, and other precinct characteristics.

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.8     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.1
## v readr   2.1.2     v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
progresa <- read.csv("data/progresa.csv")
summarize_all(progresa, ~sum(is.na(.)))
```

```
##   treatment pri2000s t1994 avgpoverty pobtot1994 villages t2000
## 1         0        0     0          0          0        0     0
```

Question 1 [7 pts]

Consider the impact of early versus late receipt of the CCT program on voter turnout in the 2000 election. No data analysis is needed for this question.

1a [3 points]

What is the specific causal question? What are the potential outcomes of a single precinct?

1b [2 points]

For precincts receiving the CCT program early, what is their average missing counterfactual outcome?

1c

How will the average missing counterfactual outcome for the treated precincts be estimated in this study?

1d

What do the researchers hypothesize the treatment effect for this outcome will be?

Answer 1

Note: There are no missing data in any fields in the data frame. ### Answer 1a

SCQ: What is the impact of early receipt of CCT program (intervention - treatment group) relative to the late receipt of CCT program (alternative - control group) on the voter turnout for the precincts in the 2000 Mexican presidential election? The potential outcomes of a single precinct are: What the voter turnout would be for the precinct if it received early CCT program. What the voter turnout would be for the precinct if it received late CCT program.

Answer 1b

MCF: What the average voter turnout would have been for precincts which received early CCT program if they instead received late CCT program and all else remained the same.

Answer 1c

The average voter turnout for the alternate group i.e. the average voter turnout for the precincts which received late CCT program will be used as the average missing counter factual for the treated group while assuming that there is no systematic difference across the baseline covariates in both of the treatment and control groups due to randomization of treatment.

Answer 1d

The researchers hypothesized treatment effect for the outcome is: In the 2000 Mexican presidential election, the precincts which received early CCT program would have higher voter turnout relative to those precincts which received late CCT program.

Question 2 [10 pts]

2a [4 points]

Estimate the impact of early versus late receipt of the CCT program on two outcomes: voter turnout in 2000 and support for the incumbent party in 2000. Do so by comparing the average electoral outcomes in the 'treated' (Early *Progresa*) precincts versus the ones in 'control' (Late *Progresa*) precincts. Use the turnout and support rates as shares of the voting eligible population (`t2000` and `pri2000s`, respectively). Interpret your results.

2b [6 points]

Consider two pretreatment covariates, poverty level and voter turnout in the 1994 election. Are these pretreatment covariates balanced between the treatment and control groups? Use appropriate summary statistics and figures to explain your answer. Discuss the implications of the distributions of these two baseline covariates for the results you estimated in the first part of this question. Use the term *confounder* in your answer.

Answer 2

Answer 2a

```
means.t200 <- tapply(progresa$t2000, progresa$treatment, mean)
means.pri <- tapply(progresa$pri2000s, progresa$treatment, mean)
means.t200
```

```
##        0        1
## 56.42893 64.33195
```

```
means.pri
```

```
##        0        1
## 34.18906 36.11215
```

```
# treatment effect on voter turnout in 2000
means.t200[2] - means.t200[1]
```

```
##        1
## 7.903022
```

```
# treatment effect on support for incumbent party in 2000

means.pri[2] - means.pri[1]
```

```
##        1
## 1.923085
```

On average, in 2000 presidential election in Mexico, in the precincts which received early CCT program, the voter turnout is 8 percentage points higher than that of the precincts which received late CCT Program. Similarly, on average, the support for the incumbent party PRI in 2000 election is 2 percentage points higher in the precincts which received early CCT program as compared to the the precincts which received late CCT program.
Answer 2b

```
# to check whether baseline covariates turnout 1994 and avg poverty
#are balanced across the treatment arms
means.t94 <- tapply(progresa$t1994, progresa$treatment, mean)
means.t94[2] - means.t94[1]
```

```
##        1
## 2.303971
```

```
means.avgp <- tapply(progresa$avgpoverty, progresa$treatment, mean)
means.avgp[2] - means.avgp[1]
```

```
##           1
## -0.02121739
```

```
# summary stats for treatment and control groups
tapply(progresa$t1994, progresa$treatment, summary)
```

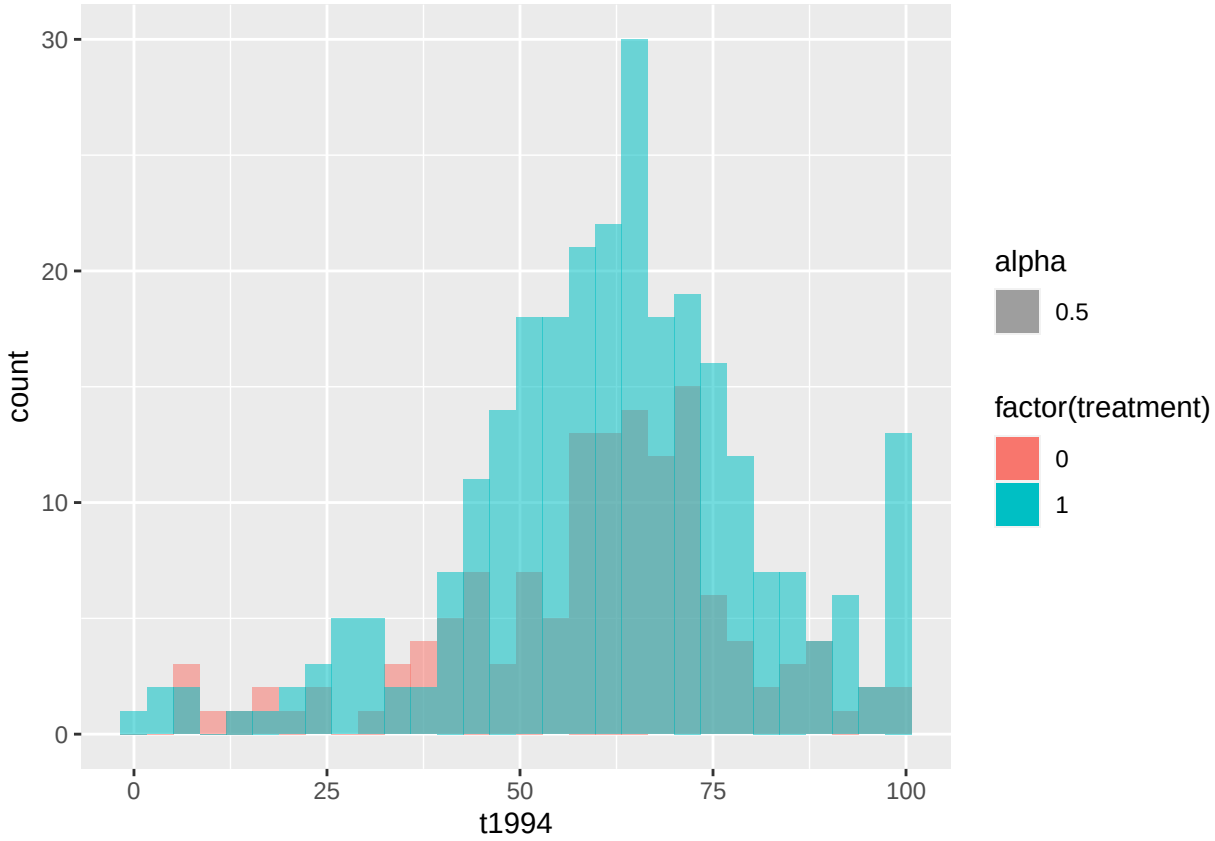
```
## $`0`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   5.308  49.913  61.309  59.554  70.970 100.000
##
## $`1`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.001  51.159  62.621  61.858  73.162 100.000
```

```
tapply(progresa$avgpoverty, progresa$treatment, summary)
```

```
## $`0`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   3.000   4.321   4.750   4.592   5.000   5.000
##
## $`1`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   3.000   4.268   4.733   4.571   5.000   5.000
```

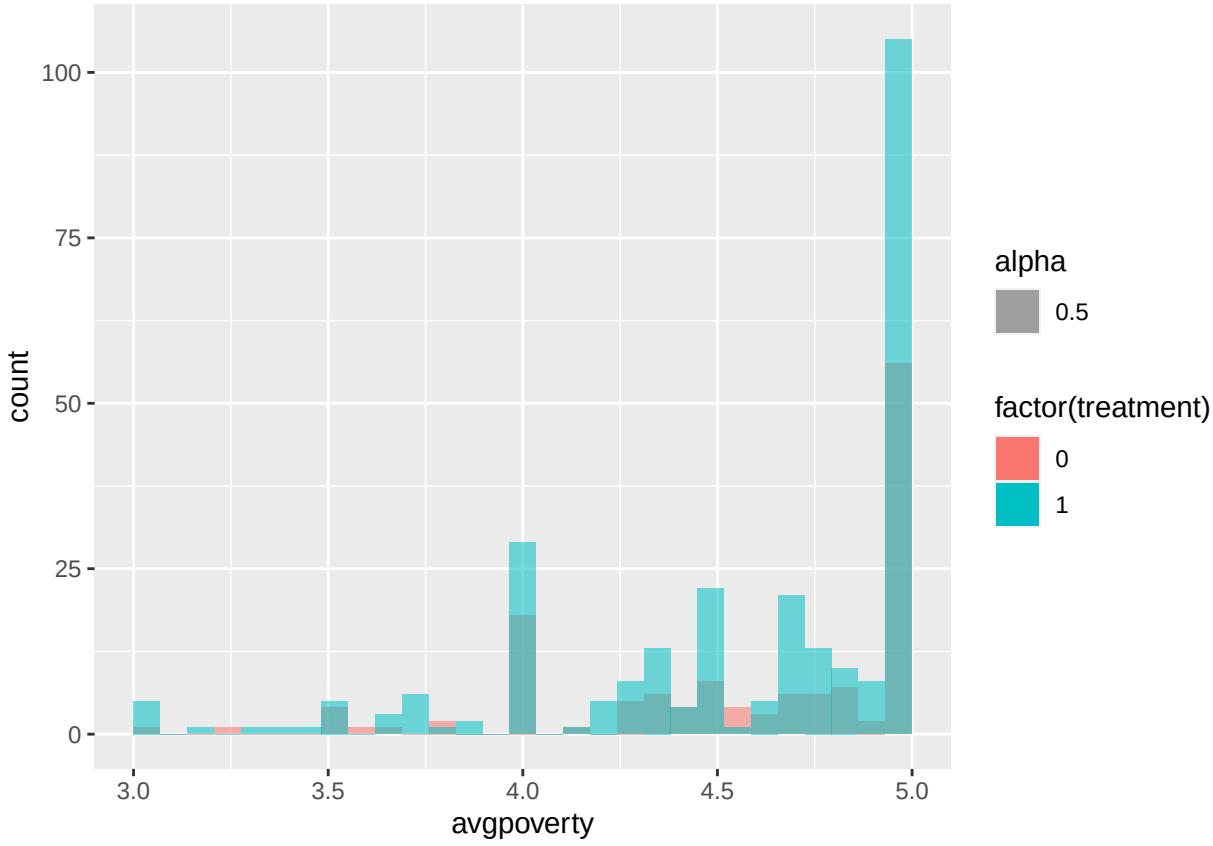
```
#overlaid histograms for the two covariates

progresa %>%
  ggplot(aes(x = t1994, fill = factor(treatment), alpha = 0.5)) +
  geom_histogram(position = 'identity')
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

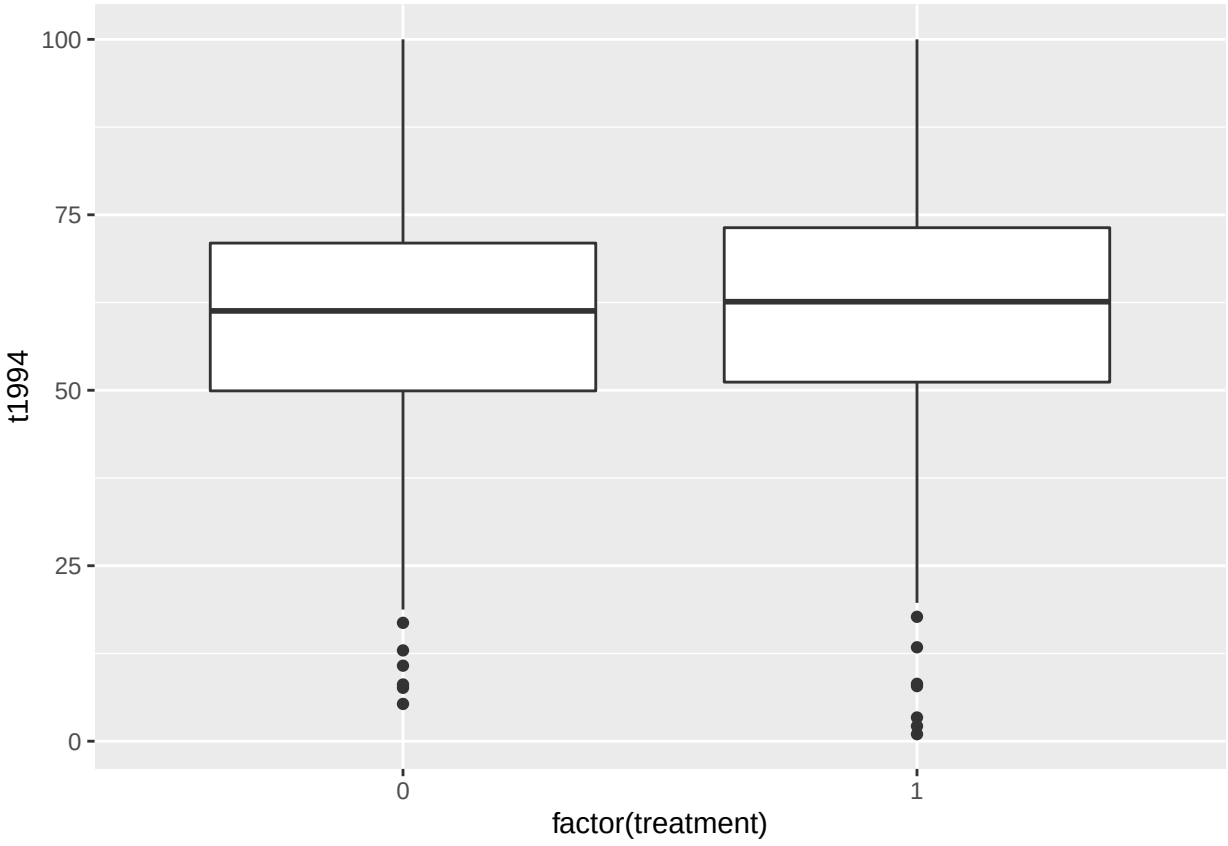



```
progresa %>%
  ggplot(aes(x = avgpoverty, fill = factor(treatment), alpha = 0.5)) +
  geom_histogram(position = 'identity')
```

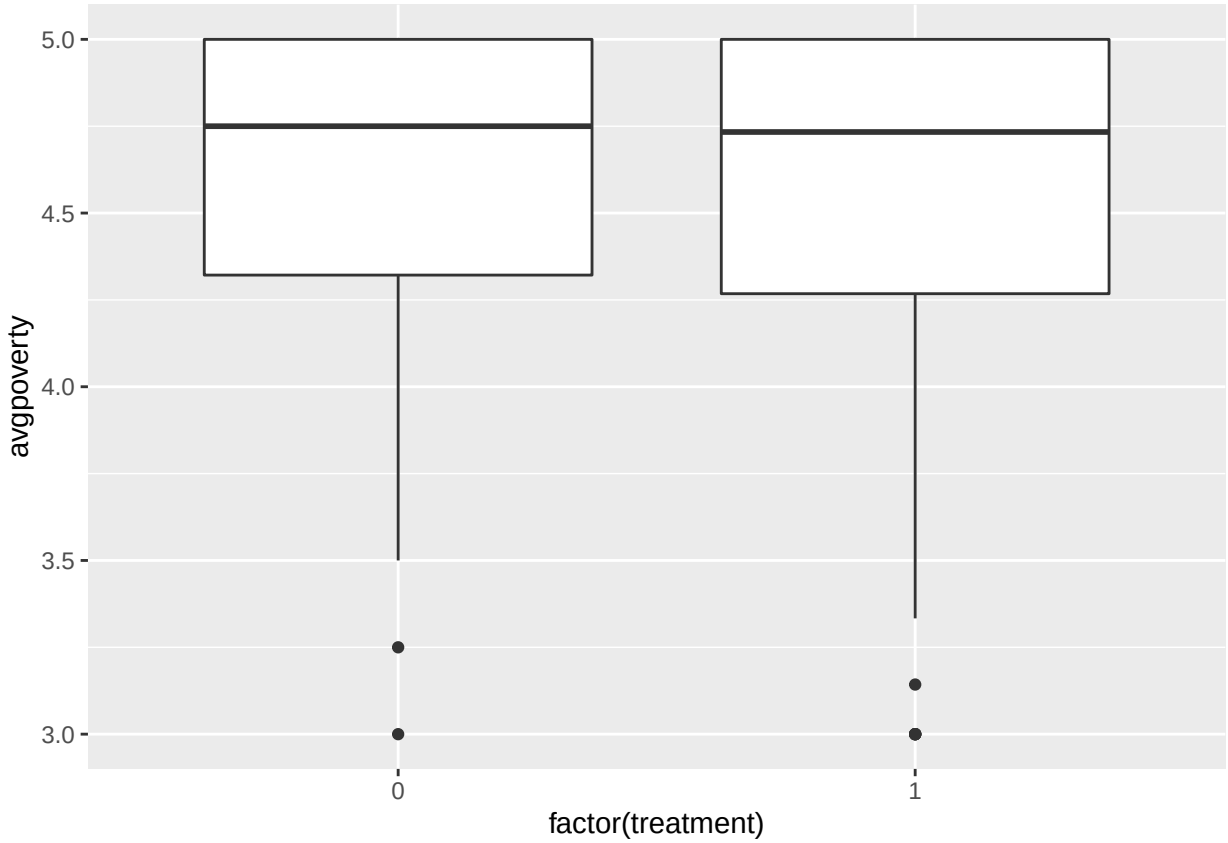
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
# side by side boxplots for the two covariates
progresa %>%
 ggplot(aes(y = t1994, x = factor(treatment))) +
  geom_boxplot()
```




```
progresa %>%
 ggplot(aes(y = avgpoverty, x = factor(treatment))) +
  geom_boxplot()
```



On average, the treatment precincts (which received early CCT program) had 2.3 percentage points higher turnout in 1994 as compared to the control group precincts (which received late CCT program). Further, on average, the treatment precincts which received early CCT program had 0.02 percentage points lesser average poverty as compared to the control group precincts which received late CCT program. Thus, the pre-treatment covariates are only slightly different. Comparison of the summary statistics of pre-treatment turnout of 1994 for the control and treatment precincts reveals that the range of turnout in 1994 in control group was between 5 and 100 percentage points while that of treatment group the range is between 1 and 100 and it shows that in the treatment group there is some outlier on the lower end of the distribution. Similarly, comparison of the average poverty level between the two groups reveal slight difference where the mean of the average poverty in the control group precincts is 4.59 percentage points while in treatment group it is around 4.57 percentage points. The overlaid histogram of turnout in 1994 between the two groups reveals that on average the turnout in the treatment group is higher than the control group and there is an overlap between the two with slight differences at the lower end. Both are unimodal and right skewed. Comparison of the poverty level in histogram shows both groups are unimodal and distribution is right skewed with slight differences at some points and there are also some gaps. The treatment group precincts are on average slightly poorer than the control group precincts. Side-by-side boxplot comparison of the turnout in 1994 between the control and the treatment group reveals that on average, the turnout is slightly lower in the control group as compared to treatment group and same trend is present in the middle 50% of data in both of the groups. Both groups have outliers on the lower end of the distribution and are right skewed. Treatment group has more outliers on the lower end which shows that there are precincts which have relatively much lower turnout and there are some gaps between the outliers. Comparison of the average poverty level between the two groups through side-by-side boxplots shows slight differences in the mean where treatment group mean is slightly lower than the control group and the first quartile of the treatment group is also lower than that of the control group which shows that the treatment group relatively comprises slightly poorer precincts. Also, there are outliers on the lower end in both of the groups with some gaps in between which shows that in both of the groups, some precincts are much poorer. Overall, there are slight differences between the two groups in terms

of 1994 turnout and average poverty. On the basis of these slight differences in the baseline covariates of average poverty and 1994 turnout, it can be concluded that both these baseline covariates are not confounders as these are not systematically different across the control and treatment groups because of the randomization of the assignment of eligible villages in the precincts to either of the two treatment arms.

Question 3 [10 pts]

Considering all precincts in the study, investigate the linear relationship between a precinct's voter turnout in the 2000 election (outcome) and it's voter turnout in the 1994 election (predictor) by answering the questions below.

3a [6 points]

Use a scatterplot and linear regression to investigate this relationship. Run the simple linear regression model. Make a scatterplot and add the estimated linear regression line to this figure. Make a residual plot and add a horizontal zero line to this figure. What do the scatterplot and residual plot tell us about the shape of this bivariate relationship? In particular, is the linearity assumption for linear regression violated or does it appear to hold? What do you see in the figure that tells you that it does or does not hold?

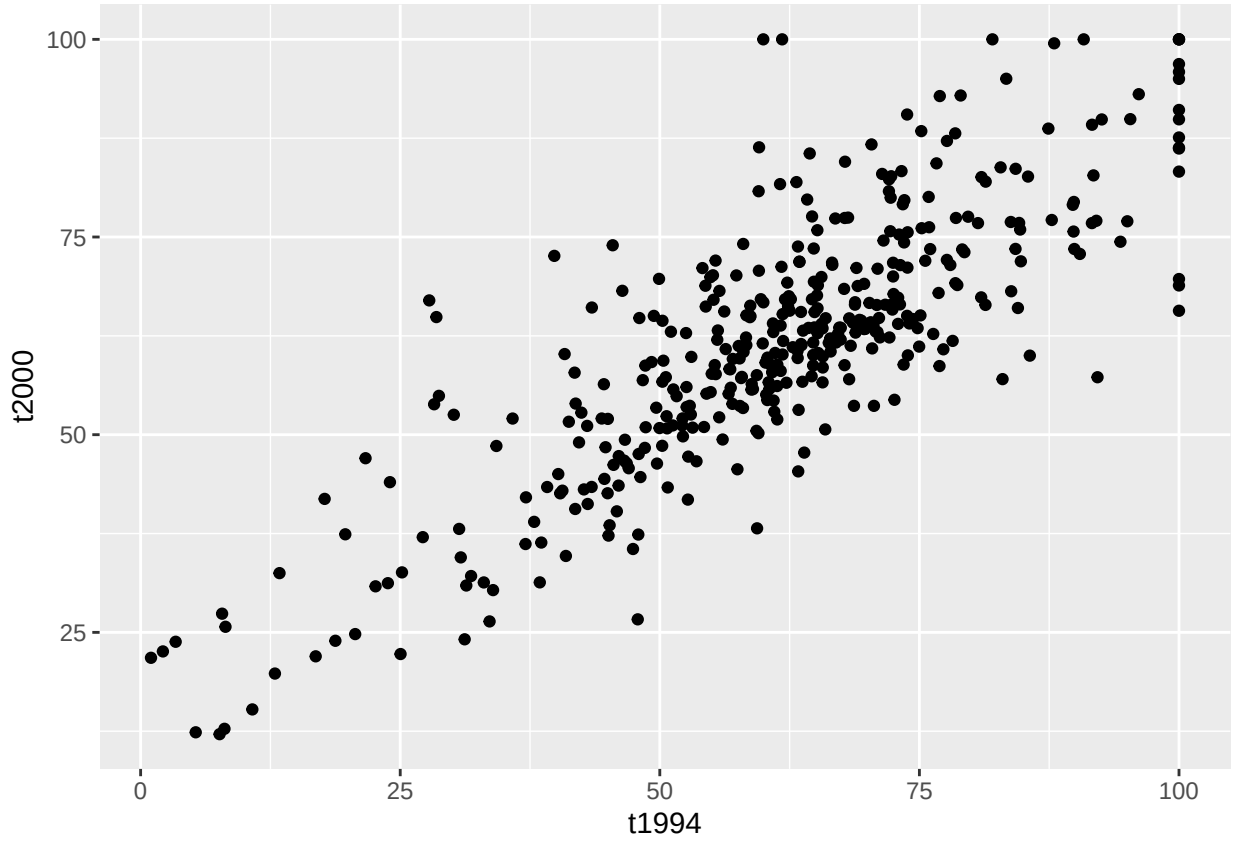
3b [4 points]

Regardless of your answer to 3a, interpret the estimated coefficients of the simple linear regression. Interpret the RMSE and R^2 value from the model.

Answer 3

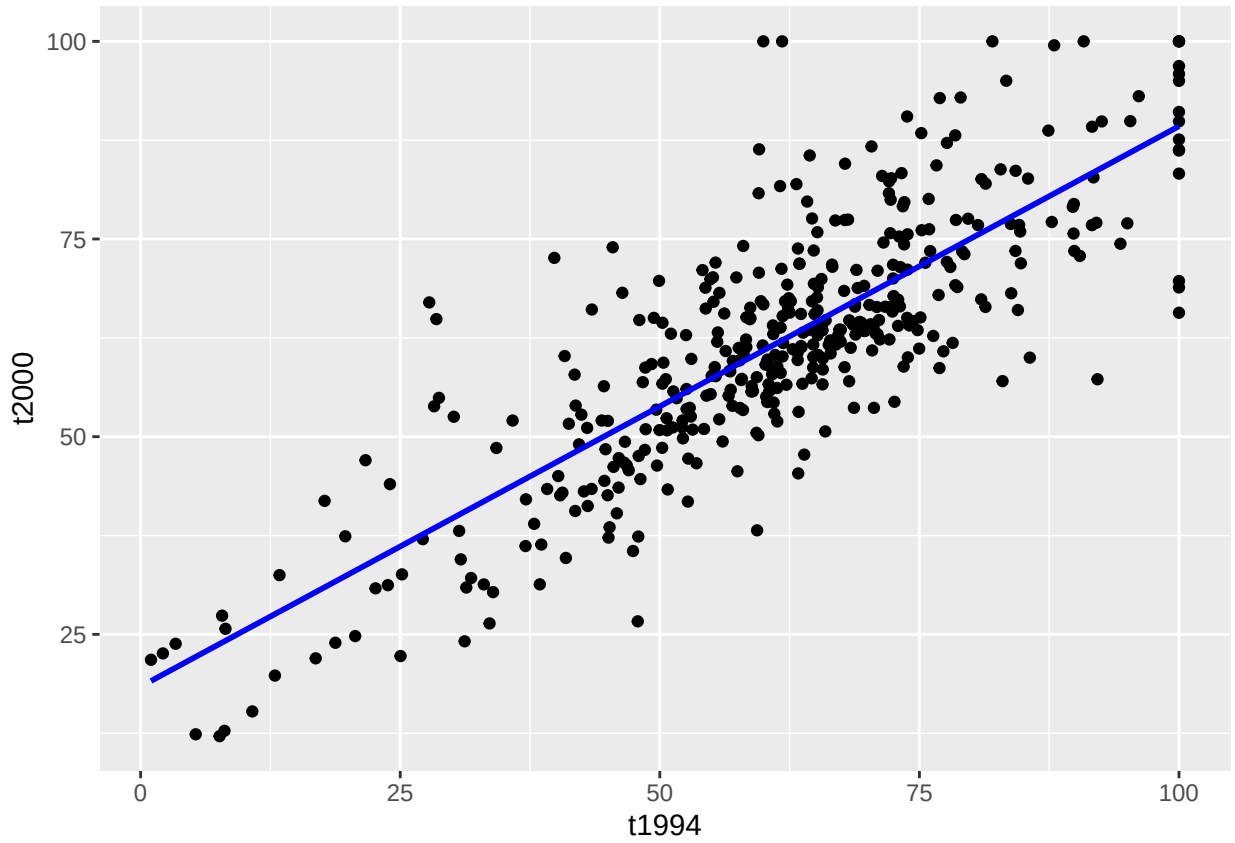
Answer 3a

```
#scatter plot
progresa %>%
  ggplot(aes(x = t1994, y = t2000)) +
  geom_point()
```




```
#adding an estimated linear regression line to scatter plot
progresa %>%
  ggplot(aes(x = t1994, y = t2000)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "blue")
```

```
## `geom_smooth()` using formula 'y ~ x'
```

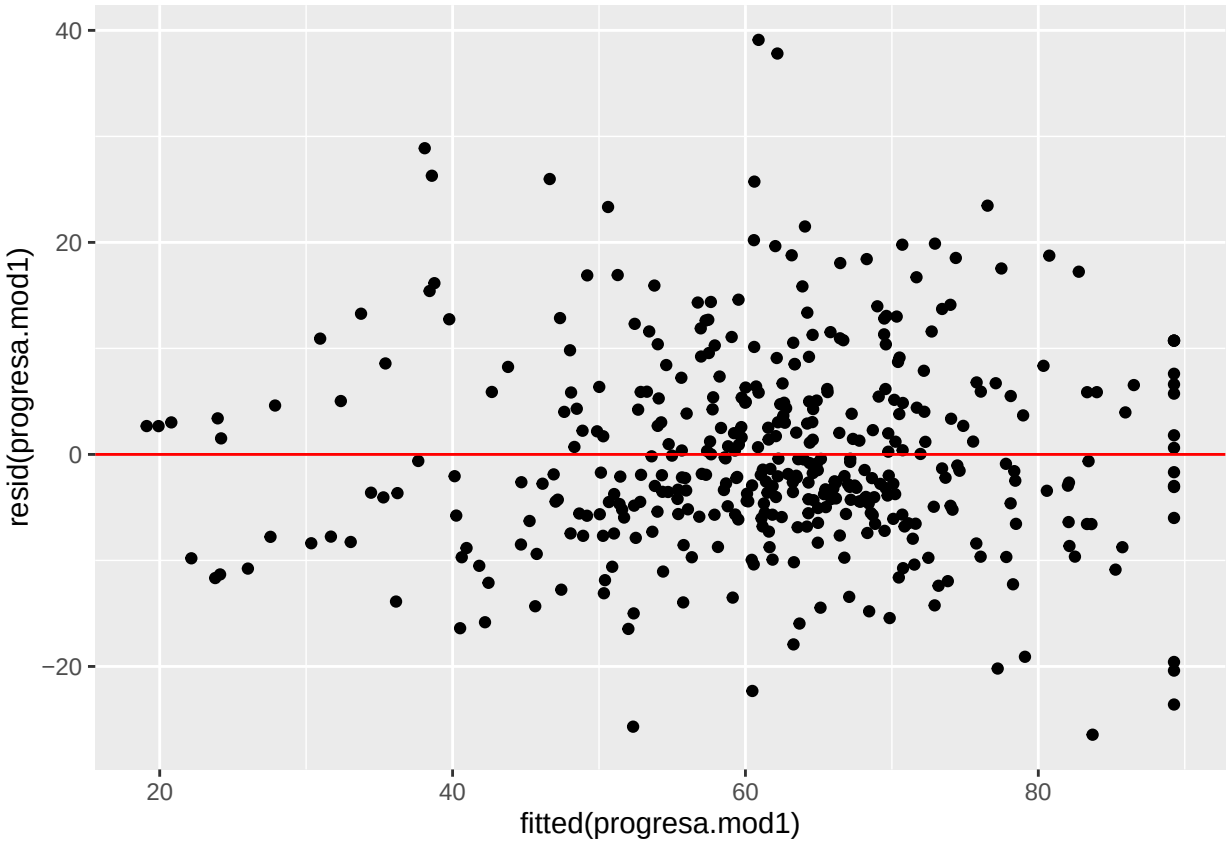



```
#linear regression
progresa.mod1 <- lm(t2000 ~ t1994 , data = progresa)
summary(progresa.mod1)
```

```
##
## Call:
## lm(formula = t2000 ~ t1994, data = progresa)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -26.439  -5.683  -1.848   5.318  39.104
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 18.39949    1.56737   11.74   <2e-16 ***
## t1994        0.70867    0.02449   28.93   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.419 on 405 degrees of freedom
## Multiple R-squared:  0.674,  Adjusted R-squared:  0.6732
## F-statistic: 837.2 on 1 and 405 DF,  p-value: < 2.2e-16
```

```
#residual polot and adding horizontal zero line.
progresa %>%
```

```
ggplot(aes(x = fitted(progresa.mod1), y = resid(progresa.mod1))) +
geom_point() +
geom_hline(yintercept = 0, color = "red")
```



Scatter plot clearly shows a medium positive association between turnout in 2000 and the turnout in 1994. Further, on the scatter plot, strong positive association is present between the respective turnouts in 1994 and 2000 in precincts where the turnouts are falling in the middle 37 to 75 percentage points as compared to precincts where turnouts are falling below 37 and above 75 percentage points. The mean of the residuals appears to be approximately zero in each segment of the residual plot, suggesting that the linearity assumption holds. Overall, both the scatter plot and residual plot help us conclude that there is medium positive linear association between turnout in 2000 and the turnout in 1994.

Answer 3b

Intercept 18.4: The y-intercept is the expected turnout in 2000 in precincts where average turnout in 1994 was 0. Zero turnout in 1994 is outside the support of data and hence the y-intercept has no practical interpretation. Slope 0.71: With a one percentage point increase in turnout in 1994, on average, we expect an increase of 0.71 percentage points in the turnout in 2000. Equivalently, voter turnout in 2000, on average, increases by 7.1 percentage points for an increase of 10 percentage points in the voter turnout in 1994. RMSE: On average, the turnout in 2000 will differ by 9.4 percentage points from what we would expect based on the 1994 turnout. R-squared: 0.674: 67.4% of the uncertainty in the turnout in 2000 is accounted for by the 1994 turnout and the linear relationship between 2000 and 1994 election turnouts.

Question 4 [22 pts]

4a [4 points]

Estimate the impact of early versus late receipt of the CCT program on voter turnout using multiple linear regression. Include two predictors in your model: *treatment*, and turnout in the 1994 election. Create a residual plot and use it to assess the linearity assumption.

4b [8 points]

Write out the multiple regression equation for this model first as a single equation and then as a pair of equations, one for each treatment arm. Create a scatterplot of the outcome and the continuous predictor variable. Color the points on this scatterplot by their treatment status. Add the two regression lines to this figure. Or sketch the regression lines described here by hand, take a picture and include it in your HW document.

4c [10 points]

Interpret the model coefficients for this multiple regression equation (the three from the single regression equation). Interpret the RMSE and the R^2 value for the model. Compare them to the RMSE and R^2 for the model in Question 3. What does this model tell you about whether the timing of the CCT program had the hypothesized effect?

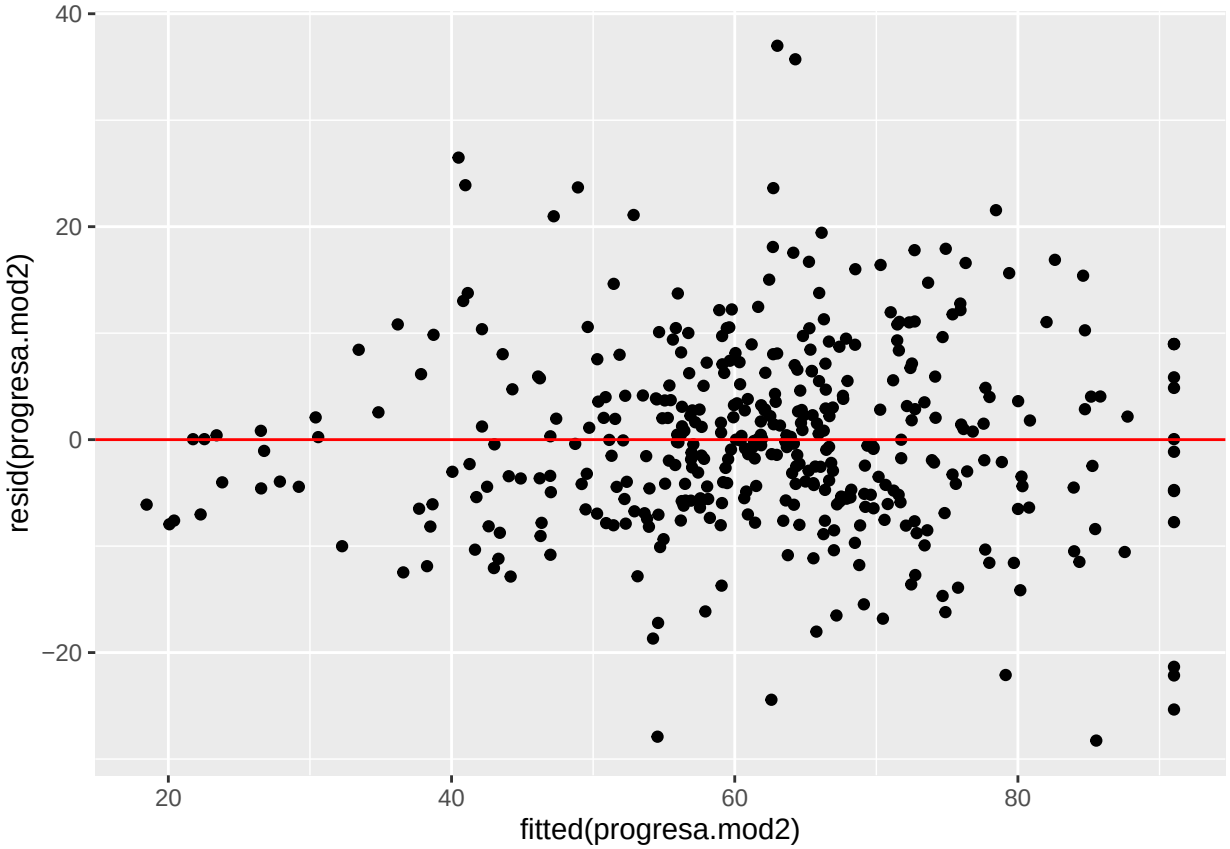
Answer 4

Answer 4a

```
progresa.mod2 <- lm(data = progresa, t2000 ~ treatment + t1994)
summary(progresa.mod2)
```

```
##
## Call:
## lm(formula = t2000 ~ treatment + t1994, data = progresa)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -28.265  -5.577  -0.374   4.792  36.992
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 14.75296    1.58622   9.301  < 2e-16 ***
## treatment    6.29070    0.94203   6.678 8.04e-11 ***
## t1994        0.69980    0.02331  30.021  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.95 on 404 degrees of freedom
## Multiple R-squared:  0.7064, Adjusted R-squared:  0.7049
## F-statistic: 485.9 on 2 and 404 DF,  p-value: < 2.2e-16
```

```
progresa %>%
  ggplot(aes(x = fitted(progresa.mod2), y = resid(progresa.mod2))) +
  geom_point() +
  geom_hline(yintercept = 0, color = "red")
```



The linearity assumption appears to hold. Our evidence for this conclusion is that the mean of the residuals appears to be zero in all segments of the residual plot as we move from left to right.

Answer 4b

$$t2000_i = 14.8 + 6.3(treatment_i) + 0.7(t1994_i) + \hat{\epsilon}_i$$

TREATED - EARLY CCT

$$t2000_i = 21 + 0.7(t1994_i) + \hat{\epsilon}_i$$

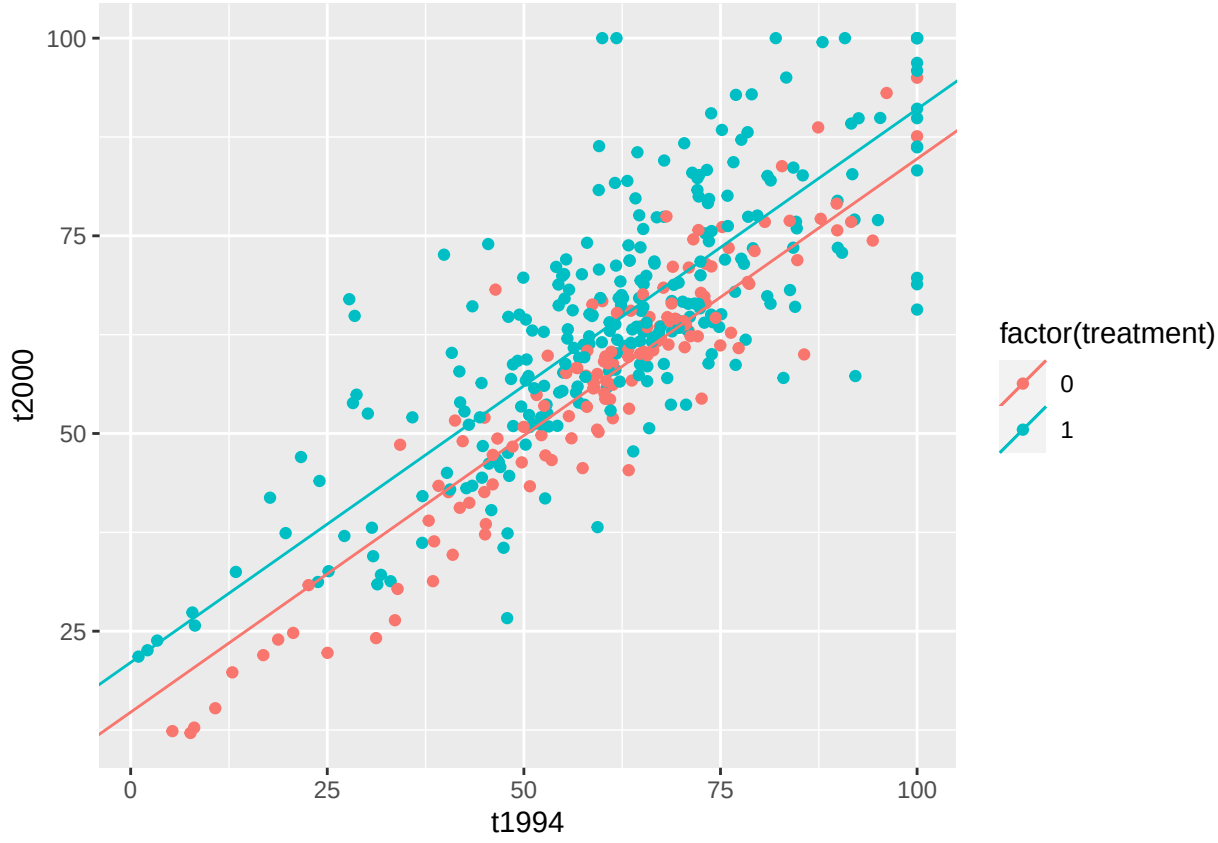
CONTROL - LATE CCT

$$t2000_i = 14.8 + 0.7(t1994_i) + \hat{\epsilon}_i$$

```
progresa.mod2$coef
```

```
## (Intercept)   treatment       t1994
##  14.7529579   6.2907049   0.6997991
```

```
#since late CCT program precincts are coded as 0, the regression
#model intercept is the y-intercept for the late CCT line
int_late = progresa.mod2$coef['(Intercept)']

#since early CCT program precincts are coded as 1, the intercept for the early
# CCT precincts line is the regression intercept PLUS the coefficient on the
# dichotomous variable
int_early = progresa.mod2$coef['(Intercept)'] + progresa.mod2$coef['treatment']

#since there is no interaction effect, both lines have the same slope
slope = progresa.mod2$coef['t1994']  ##coefficient on the continuous variable

progresa %>%
  ggplot(aes(y = t2000, x = t1994, color = factor(treatment))) +
  geom_point() +
  geom_abline(aes(intercept = int_late, slope = slope, color = '0')) +  #regression line for late CCT p
  geom_abline(aes(intercept = int_early, slope = slope, color = '1'))  #regression line for early CCT p
```



Answer 4c

Y-intercept 14.8: The expected voter turnout in 2000 is 14.8 percentage points for the precincts where 1994 turnout was 0 and which received late CCT program (control group). It is mathematically necessary only and is beyond the support of data. Coefficients: 1) Turnout in 2000 increases on average by 0.7 percentage points for a one percentage point increase in 1994 turnout within each treatment arm. Equivalently, turnout in 2000 increases on average by 7 percentage points for a 10 percentage point increase in 1994 turnout within

each treatment arm. 2) Holding turnout in 1994 constant, the treatment of early CCT program increases turnout in 2000 by 6.3 percentage points on average. RMSE 8.95: On average, the voter turnout in 2000 is 8.95 percentage points away from the expected turnout based on 1994 turnout and the treatment status. Multiple R-squared: 0.7064: 70.64% of the uncertainty in the turnout in 2000 is accounted for by the two predictors of 1994 turnout and treatment arm (early and late CCT program) and the linear relationship of these variables with the 2000 election turnout. This model tells us that the timing of the CCT program does have the hypothesized effect as the precincts which received early CCT program have on average 6.3 percentage point higher turnout in 2000.

Question 5 [19 pts]

Now, we will explore whether early versus late receipt of the CCT program affects 2000 voter turnout differently for precincts that had low versus high voter turnout in the prior 1994 election.

5a [2 points]

Add an interaction term to your model from Question 4 between 1994 voter turnout and the treatment variable.

5b [9 points]

Write out the multiple regression equation for this model first as a single equation and then as a pair of equations, one for each treatment arm. Create a scatterplot of the outcome and the continuous predictor variable. Color the points on this scatterplot by their treatment status. Add the two regression lines to this figure. Or sketch the regression lines described here by hand, take a picture and include it in your HW document.

5c [8 points]

Interpret all four of the model coefficients for this multiple regression equation, or interpret the four model coefficients in the two separate simple linear regression equations. Interpret the RMSE and the R^2 value for the model. Compare them to the RMSE and R^2 for the model in Question 4. What does this model tell you about whether the timing of the CCT program had more or less of an effect on precincts with prior low voter turnout rates relative to precincts with prior high voter turnout rates?

Answer 5

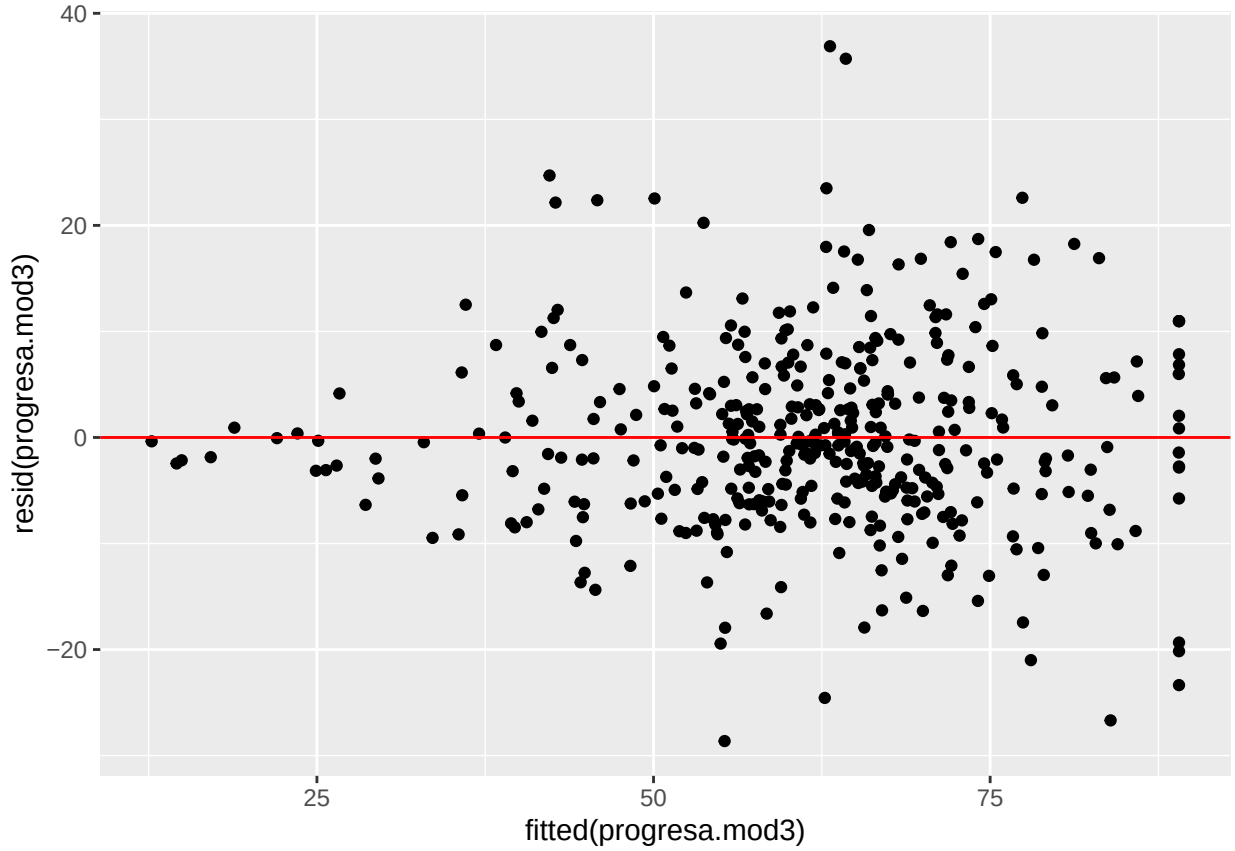
Answer 5a

```
progresa.mod3 <- lm(data = progresa, t2000 ~ treatment + t1994 + t1994*treatment)
summary(progresa.mod3)
```

```
##
## Call:
## lm(formula = t2000 ~ treatment + t1994 + t1994 * treatment, data = progresa)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -28.632  -5.396  -0.741   4.704  36.893
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       8.44778    2.50423   3.373 0.000814 ***
## treatment        15.83088    3.09763   5.111 4.97e-07 ***
## t1994             0.80567    0.04007  20.105  < 2e-16 ***
## treatment:t1994 -0.15817    0.04898  -3.229 0.001343 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 8.847 on 403 degrees of freedom
## Multiple R-squared:  0.7138, Adjusted R-squared:  0.7116
## F-statistic:   335 on 3 and 403 DF,  p-value: < 2.2e-16
```

```
progresa %>%
  ggplot(aes(x = fitted(progresa.mod3), y = resid(progresa.mod3))) +
  geom_point() +
  geom_hline(yintercept = 0, color = "red")
```



Answer 5b

$$t2000_i = 8.5 + 16(treatment_i) + 0.8(t1994_i) - 0.2(treatment_i) * (t1994_i) + \hat{\epsilon}_i$$

TREATED - EARLY CCT

$$t2000_i = 24.5 + 0.6(t1994_i) + \hat{\epsilon}_i$$

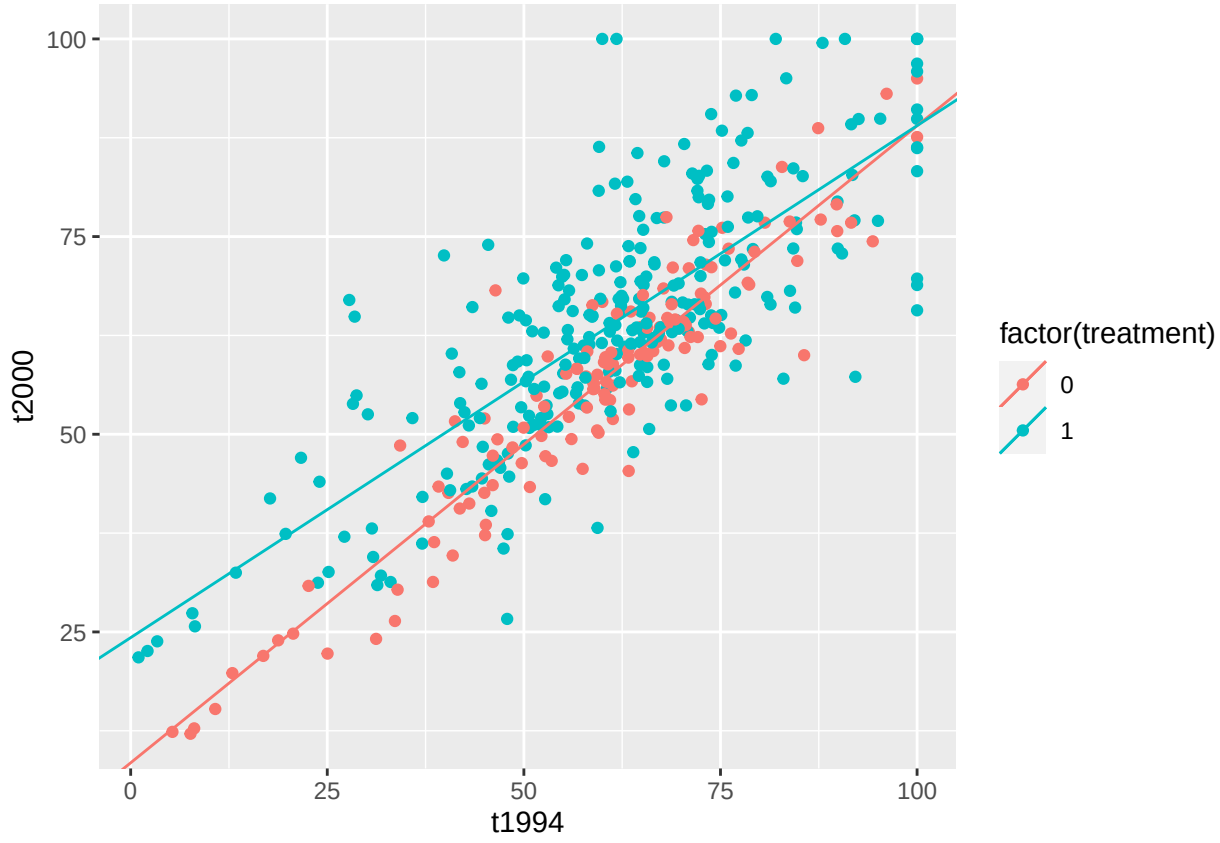
CONTROL - LATE CCT

$$t2000_i = 8.5 + 0.8(t1994_i) + \hat{\epsilon}_i$$

```
progresa.mod3$coef
```

```
##     (Intercept)       treatment          t1994 treatment:t1994
##       8.4477812      15.8308777       0.8056721      -0.1581699
```

```r
#since late CCT program precincts are coded as 0,
#the regression model intercept is the y-intercept for the late CCT line
int_late1 = progresa.mod3$coef['(Intercept)']
#the slope for the late CCT program line is just the regression model #coefficient on the continuous va
slope_late = progresa.mod3$coef['t1994']

#since early CCT program precincts are coded as 1, the intercept for the early
# CCT precincts line is the regression intercept PLUS the coefficient on the
# dichotomous variable
int_early1 = progresa.mod3$coef['(Intercept)'] + progresa.mod3$coef['treatment']

#the slope for the early CCT program precincts line is the coefficient on the #continuous variable of t
slope_interact = progresa.mod3$coef['t1994'] + progresa.mod3$coef['treatment:t1994']

progresa %>%
  ggplot(aes(y = t2000, x = t1994, color = factor(treatment))) +
  geom_point() +
  geom_abline(aes(intercept = int_late1, slope = slope_late, color = '0')) +  #regression line for late
  geom_abline(aes(intercept = int_early1, slope = slope_interact, color = '1')) #regression line for ea
```



Answer 5c

Control Group (precincts receiving late CCT program): Y intercept 8.44778: The expected turnout in 2000 is 8.4 percentage points for precincts which are in the control group (received late CCT program) and where turnout in 1994 was 0 percentage points. It is not meaningful and beyond support of data.

Coefficient (t1994 0.80567): In the control group, on average, with a one percentage point increase in the turnout in 1994, the expected voter turnout in 2000 increases by 0.81 percentage points. Equivalently, on average, with a 10 percentage point increase in the turnout in 1994, the expected voter turnout in 2000 will increase by 8.1 percentage points in this group.

Treatment Group (precincts receiving early CCT program): Y Intercept 24.5: The expected turnout in 2000 is 24.5 percentage points for precincts which are in the treatment group (received early CCT program) and where turnout in 1994 was 0 percentage points. It is not meaningful and beyond support of data.

Coefficient (t1994 0.6): In the treatment group, on average, with a unit percentage point increase in the turnout in 1994, the expected voter turnout in 2000 increases by 0.6 percentage point. Equivalently, on average, with a 10 percentage point increase in the turnout in 1994, the expected voter turnout in 2000 will increase by 6 percentage points in this group. Residual standard error: 8.85 RMSE 8.85: On average, the voter turnout in 2000 is 8.85 percentage points away from the expected turnout based on 1994 turnout, the treatment status, and interaction between 1994 turnout and the treatment. Multiple R-squared: 0.7138 71.38% of the uncertainty in the turnout in 2000 (outcome) is accounted for by the predictors 1994 turnout, treatment arm, and their interaction to predict 2000 turnout in a multiple linear regression. While comparing the RMSE and R^2 of this multiple linear model with interaction (which are 8.85 and 71.38 respectively) with the multiple linear regression without interaction (where these are 8.95 and 70.64 respectively), we see that in this model RMSE has gone down from 8.95 in the previous model of without interaction to 8.85 in this model with interaction whereas the R^2 has improved and has gone up from 70.64 in previous model to 71.38 in this model and with the reduced error and more certainty this model has got better for prediction as compared to the last one and reduces 71.38% uncertainty. Further, this model tells that timing of the CCT program had more effect on precincts with lover turnout rates relative to precincts with prior high voter turnout rates. Also, overall, the treatment of early CCT program has caused higher turnout in 2000 all the times compared to the 1994 turnout in precincts with late CCT program.