Homework :Chengkan Tao

chengkan_tao

```
devtools::install_github(c('rstudio/rmarkdown', 'yihui/tinytex'))
tinytex::install_tinytex()
```

#1

#estimating

There are two models:one is LPM with quadratic for some variables , the
other is logit model with cube for some variables.

```
##
## Call:
## lm(formula = treat ~ age + agesq + educ + educsq + marr + nodegree +
##     black + hisp + re74 + re75 + u74 + u75 + interaction2, data = nsw_dw_cpscontrol)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.09366 -0.00605  0.00137  0.00485  1.00547
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -4.699e-02  7.320e-03  -6.419 1.39e-10 ***
## age           1.798e-03  3.083e-04   5.831 5.56e-09 ***
## agesq        -2.686e-05  4.178e-06  -6.430 1.30e-10 ***
## educ          3.330e-03  7.357e-04   4.527 6.01e-06 ***
## educsq       -1.371e-04  2.937e-05  -4.669 3.04e-06 ***
## marr         -9.477e-03  1.101e-03  -8.604  < 2e-16 ***
```

```
## nodegree      7.988e-03  1.340e-03   5.963 2.50e-09 ***
## black         5.106e-02  1.500e-03  34.041  < 2e-16 ***
## hisp          4.566e-03  1.735e-03   2.632  0.00850 **
## re74          2.521e-07  9.482e-08   2.659  0.00784 **
## re75         -6.582e-08  9.485e-08  -0.694  0.48768
## u74           2.142e-02  1.737e-03  12.331  < 2e-16 ***
## u75           1.007e-02  1.765e-03   5.704 1.18e-08 ***
## interaction2 -8.617e-03  4.702e-03  -1.832  0.06689 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07294 on 32415 degrees of freedom
## Multiple R-squared:  0.06258,    Adjusted R-squared:  0.0622
## F-statistic: 166.4 on 13 and 32415 DF,  p-value: < 2.2e-16


##
## Call:
## glm(formula = treat ~ age + agesq + agecube + educ + educsq +
##     educcube + marr + nodegree + black + hisp + re74 + re75 +
##     u74 + u75 + interaction1, family = binomial(link = "logit"),
##     data = nsw_dw_cpscontrol)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.2386  -0.0429  -0.0164  -0.0071   3.7553
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.353e+01  3.326e+00  -7.074 1.50e-12 ***
## age           1.512e+00  2.977e-01   5.080 3.77e-07 ***
## agesq        -4.164e-02  9.589e-03  -4.342 1.41e-05 ***
## agecube       3.468e-04  9.731e-05   3.564 0.000366 ***
## educ          3.045e-01  5.120e-01   0.595 0.552091
## educsq       -4.952e-03  5.723e-02  -0.087 0.931046
```

```
## educcube    -1.132e-03  2.006e-03  -0.564 0.572585
## marr        -1.042e+00  2.208e-01  -4.717 2.40e-06 ***
## nodegree     3.805e-01  2.875e-01   1.323 0.185683
## black        3.440e+00  2.621e-01  13.127  < 2e-16 ***
## hisp         1.657e+00  3.917e-01   4.231 2.33e-05 ***
## re74        -9.821e-05  9.154e-05  -1.073 0.283339
## re75        -1.766e-04  3.475e-05  -5.083 3.72e-07 ***
## u74          1.252e+00  2.609e-01   4.798 1.60e-06 ***
## u75         -2.427e-01  2.243e-01  -1.082 0.279213
## interaction1 1.040e-05  7.843e-06   1.326 0.184712
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2280.5  on 32428  degrees of freedom
## Residual deviance: 1198.1  on 32413  degrees of freedom
## AIC: 1230.1
##
## Number of Fisher Scoring iterations: 11
```
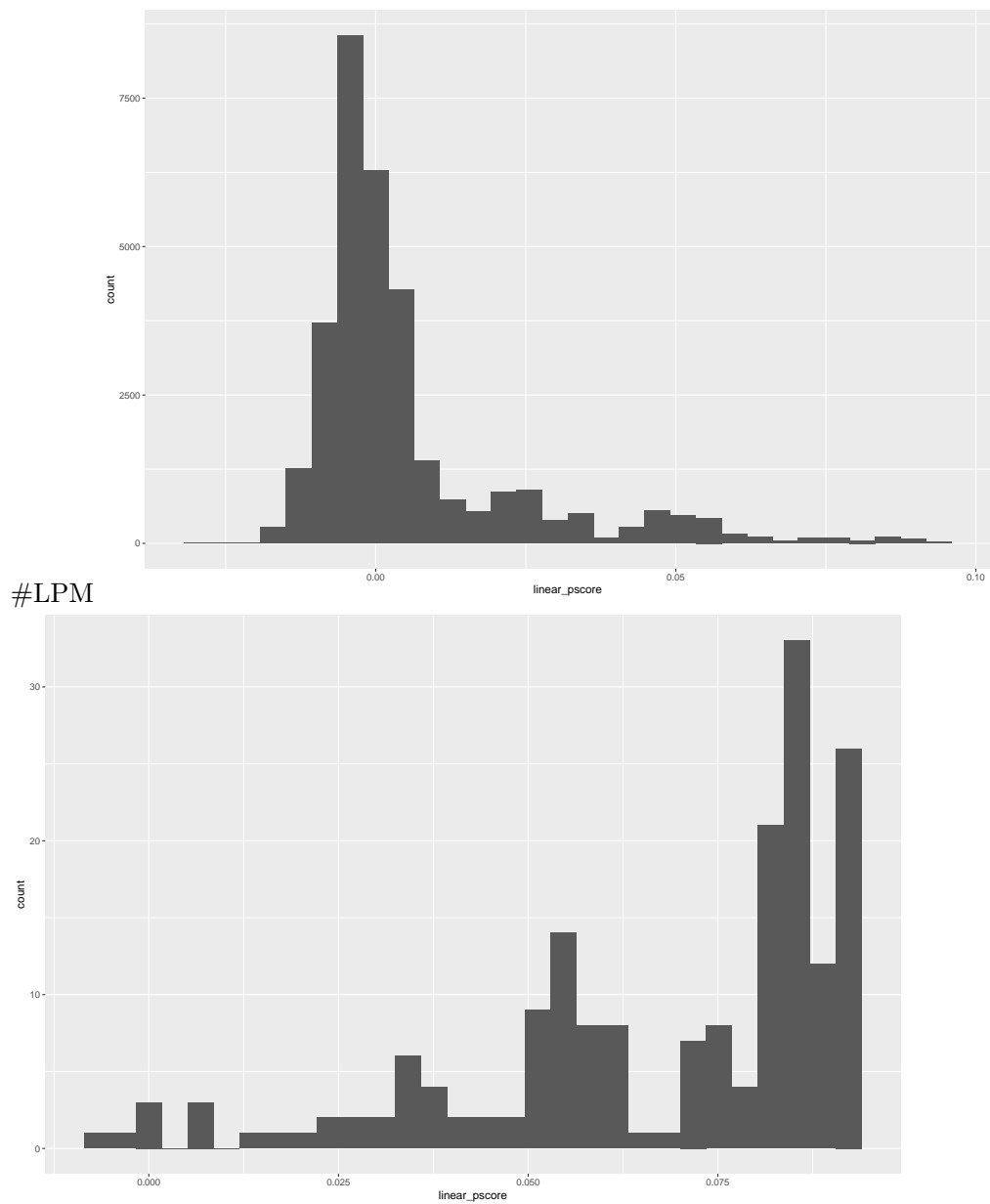
We add score into dataset
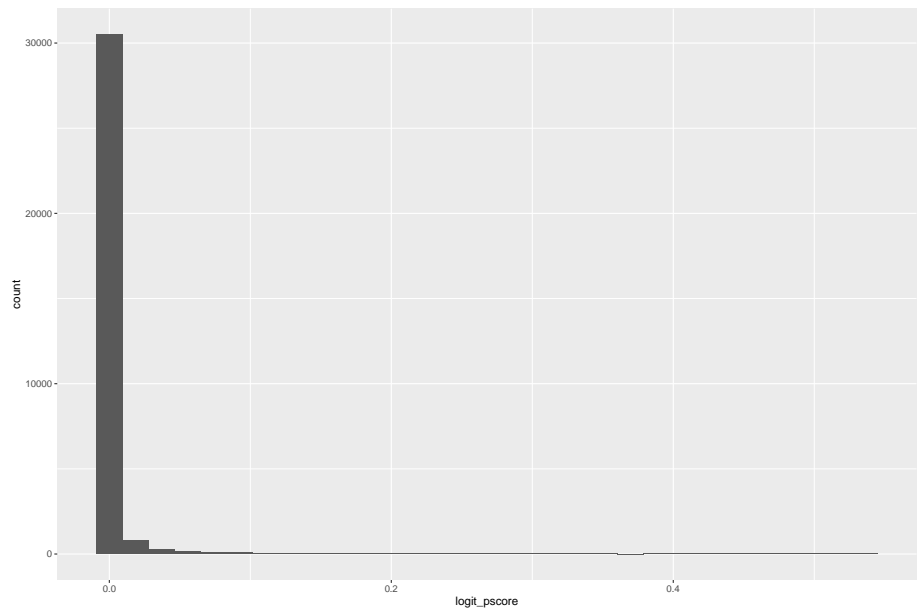
#histogram #logit
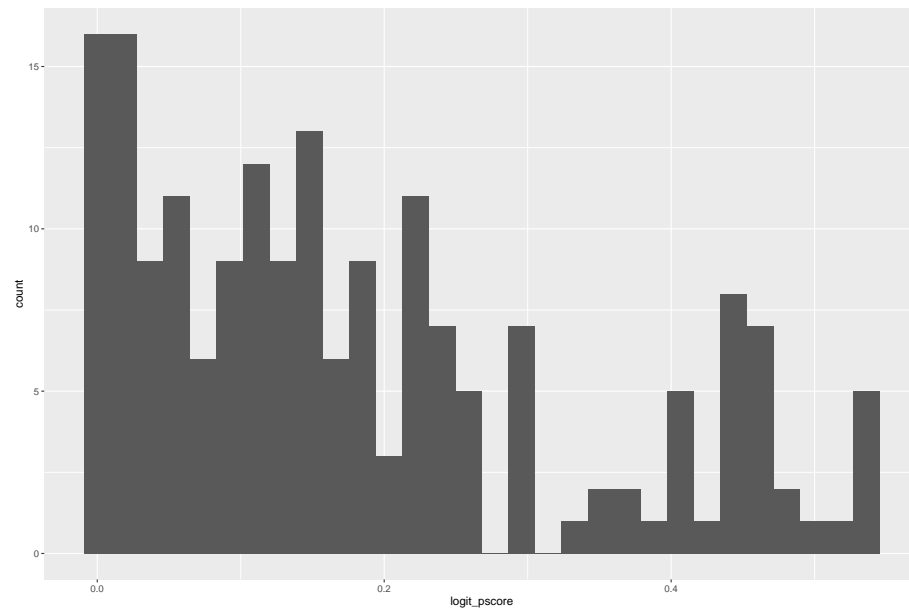
The first histogram doesn't look good

#LPM

Four histogram shows the PS distribution of two models pretty well.

#min and max

[1] 0.5356242

```
## [1] 1.122391e-07
```

This is min and max of propensity score of logit model in the control group

```
## [1] 0.5369822
```

```
## [1] 0.0008663789
```

This is min and max of propensity score of logit model in the treatment group. The max in this group is close to max in the control group. Both min is very close to zero.

```
## [1] 0.09365895
```

```
## [1] -0.0300561
```

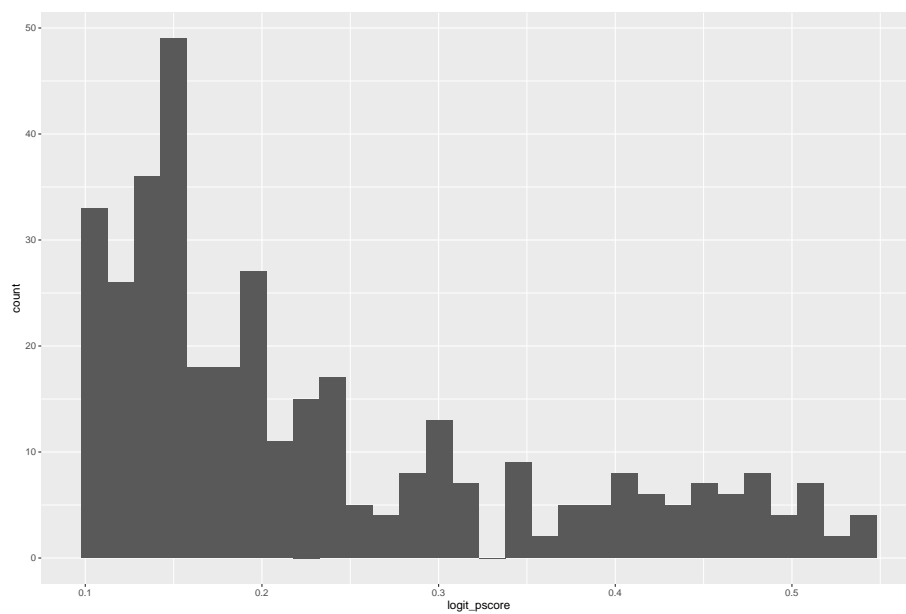This is min and max of propensity score of LPM in the control group

```
## [1] 0.09360747
```
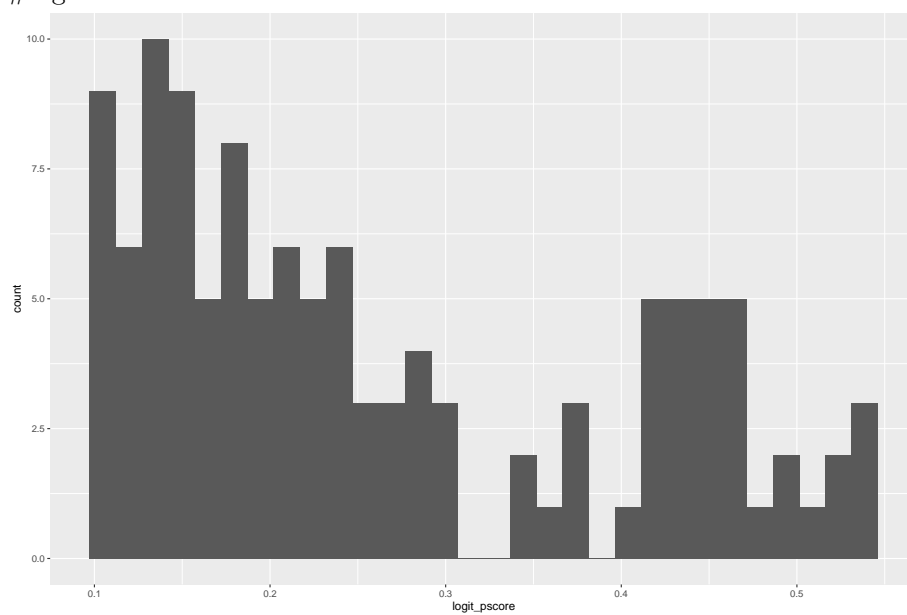
```
## [1] -0.005473957
```

This is min and max of propensity score of LPM in the treatment group. Both max and min are close to min and max in the control group.
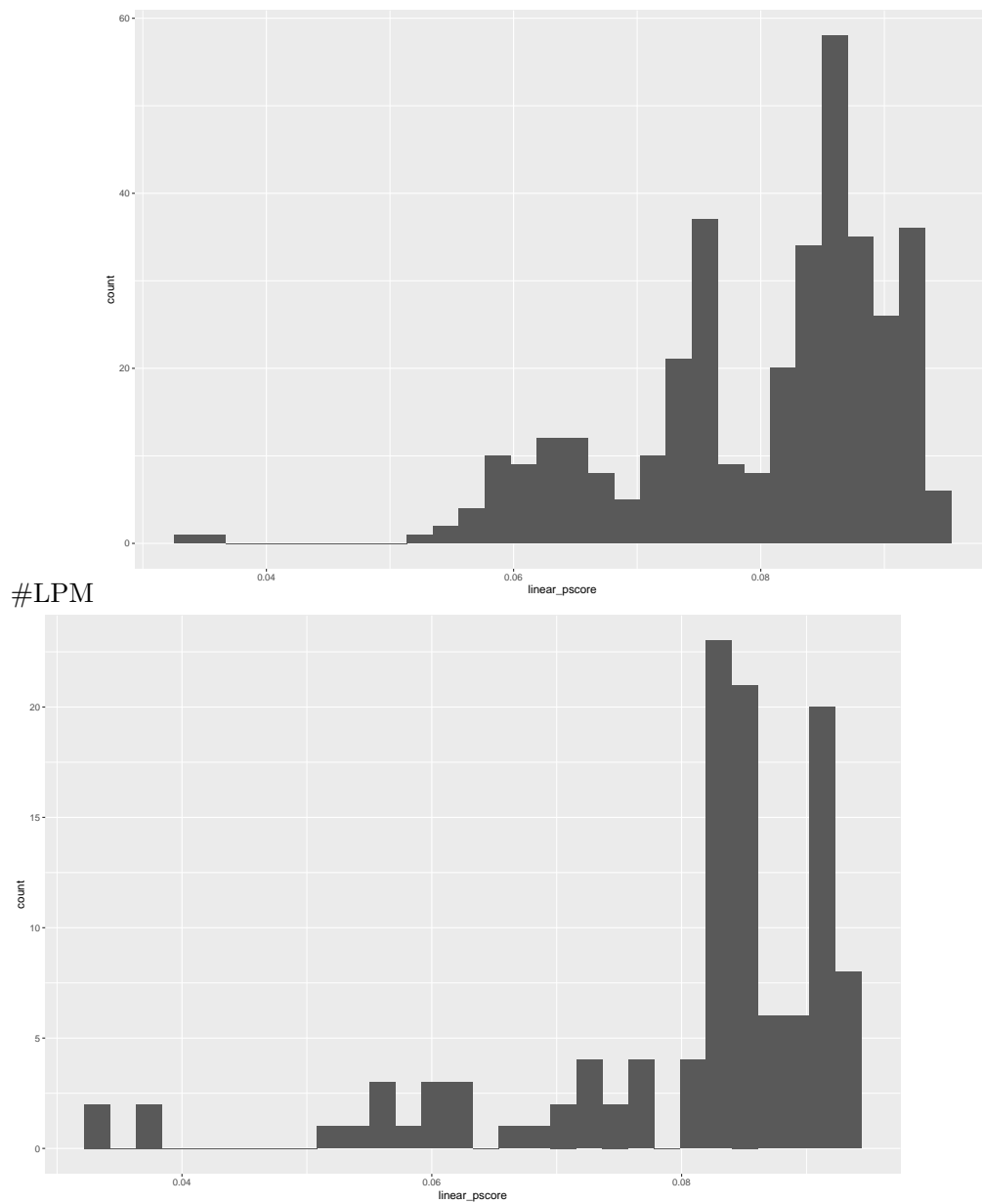
#drop

Drop all units whose propensity scores are less than 0.1 and more than 0.9

#logit

#LPM

These four histogram shows the PS distribution of two models. The distributions in the same model look similar.

#min and max

```
## [1] 0.5356242
```

```
## [1] 0.1002543
```

This is min and max of propensity score of logit model in the control group

```
## [1] 0.5369822
```

```
## [1] 0.1030246
```

This is min and max of propensity score of logit model in the treatment group.

```
## [1] 0.09365895
```

```
## [1] 0.03285054
```

This is min and max of propensity score of LPM in the control group.

```
## [1] 0.09360747
```

```
## [1] 0.03343929
```

This is min and max of propensity score of LPM in the treatment group.

We can find in the same model, min and max in control group are nearly equal to min and max in the treatment group.

#2

```
## [1] -11.45296
```

```
## [1] 265.1463
```

```
## [1] 1794.342
```

The training program caused real earnings in 1974 to decrease by \$11.453
The training program caused real earnings in 1975 to increase by \$265.146
The training program caused real earnings in 1978 to increase by \$1794.342

#3

```
## [1] -11873.59
```

```
## [1] -7592.795
```

```
## [1] -12403.43
```

```
## [1] -8545.048
```

I use two models to estimate ATT. In the logit model, i found estimated ATT of -$11873.59, and -$7592.795 with the normalization of the weights. In the LPM, i found estimated ATT of -$12403.43, and -$8545.048 with the normalization of the weights.