Project2

utkrsh

2024-05-13

Importing required libraries

```
library(tidyverse)    # ggplot(), %>%, mutate(), and friends
library(scales)       # Format numbers with functions like comma(), percent(), and dollar()
library(broom)        # Convert models to data frames
library(modelsummary) # Side-by-side regression tables
library(foreign)   # for importing stata data
library(readr)
library(haven)
library(modelsummary)
library(stargazer)
library(tidyverse)
```

Importing dataset "INJURY"

```
dataset <- read_dta("C:/Users/utkrsh/Desktop/EC402_Assignment_2_DID/INJURY.DTA")
```

Cleaning the data so that it only includes rows from (ky == 1)

```
injury <- dataset %>% filter(ky==1)
```

Renaming the columns

```
injury <- injury %>% rename(duration = durat, log_duration = ldurat,after_1980 = afchnge)
```

Viewing the dataset

```
injury
```

```
## # A tibble: 5,626 x 30
##    duration after_1980 highearn  male married  hosp indust injtype   age prewage
##       <dbl>      <dbl>    <dbl> <dbl>   <dbl> <dbl>  <dbl>   <dbl> <dbl>   <dbl>
##  1        1          1        1     1       0     1      3       1    26    405.
##  2        1          1        1     1       1     0      3       1    31    644.
##  3       84          1        1     1       1     1      3       1    37    398.
##  4        4          1        1     1       1     1      3       1    31    528.
##  5        1          1        1     1       1     0      3       1    23    529.
##  6        1          1        1     1       1     0      3       1    34    614.
##  7        7          1        1     1       1     0      3       1    35    546
##  8        2          1        1     1       1     1      3       1    45    660.
##  9      175          1        1     1       1     1      3       1    41    479.
## 10       60          1        1     1       1     1      3       1    33    481.
## # i 5,616 more rows
## # i 20 more variables: totmed <dbl>, injdes <dbl>, benefit <dbl>, ky <dbl>,
## #   mi <dbl>, log_duration <dbl>, afhigh <dbl>, lprewage <dbl>, lage <dbl>,
## #   ltotmed <dbl>, head <dbl>, neck <dbl>, upextr <dbl>, trunk <dbl>,
## #   lowback <dbl>, lowextr <dbl>, occdis <dbl>, manuf <dbl>, construc <dbl>,
## #   highlpre <dbl>
```

Converting "industry" and "injury type" to categories/factors

```
df <- injury %>% mutate(indust = as.factor(indust),injtype = as.factor(injtype))
```

1. Calculating the policy effect on duration, without running any regression, here we see the mean duration of the weeks for both treated and control group before and after 1980 (treatement)

```
difr <- df %>% group_by(after_1980,highearn) %>% summarize(mean.duration = mean(duration))
```

```
## `summarise()` has grouped output by 'after_1980'. You can override using the
## `.groups` argument.
```

```
print(difr)
```

```
## # A tibble: 4 x 3
## # Groups:   after_1980 [2]
##   after_1980 highearn mean.duration
##        <dbl>    <dbl>         <dbl>
## 1          0        0          6.27
## 2          0        1         11.2
## 3          1        0          7.04
## 4          1        1         12.9
```

[after_1980(0);highearn(0)]: pre-treatement control group mean_duration : 6.47 weeks

[after_1980(0);highearn(1)]: pre-treatement treatement group mean_duration : 11.76 weeks

[after_1980(1);highearn(0)]: post-treatement control group mean_duration : 7.03 weeks

[after_1980(1);highearn(1)]: post-treatement treatement group mean_duration : 12.89 weeks

policy_effect =[(avg_duration(post-treatement treated group)-avg_duration(pre-treatement treated group)]-[avg_duration(post-treatement control group)-avg_duration(pre-treatement control group)]

policy_effect = [avg_duration of treated group (POST-PRE)]-[avg_duration of control group(POST-PRE)]

```r
pre_treatement_treated_group <- difr %>%
  filter(after_1980 == 0, highearn == 1) %>%
  pull(mean.duration)

pre_treatement_control_group<- difr %>%
  filter(after_1980 == 0, highearn == 0) %>%
  pull(mean.duration)

post_treatement_treated_group <- difr %>%
  filter(after_1980 == 1, highearn == 1) %>%
  pull(mean.duration)

post_treatement_control_group <- difr %>%
  filter(after_1980 == 1, highearn == 0) %>%
  pull(mean.duration)
```

```r
treatement_group_before_after <- post_treatement_treated_group - pre_treatement_treated_group
control_group_before_after    <-post_treatement_control_group- pre_treatement_control_group
```

Policy Effect(DiD Estimate:)

```r
policy_effect_nive <- treatement_group_before_after - control_group_before_after
print(policy_effect_nive)
```

```
## [1] 0.9512506
```

2. Calculating the policy effect on duration, without running any regression, here we see the mean log_duration in weeks for both treated and control group before and after 1980(treatemnt)

```r
difr_log <- df %>% group_by(after_1980,highearn) %>% summarize(mean.log_duration = mean(log_duration))
```

```
## `summarise()` has grouped output by 'after_1980'. You can override using the
## `.groups` argument.
```

```r
head(difr_log)
```

```
## # A tibble: 4 x 3
## # Groups:   after_1980 [2]
##   after_1980 highearn mean.log_duration
##        <dbl>    <dbl>             <dbl>
## 1          0        0              1.13
## 2          0        1              1.38
## 3          1        0              1.13
## 4          1        1              1.58
```

[after_1980(0);highearn(0)]: pre-treatement control group mean_log_duration : 1.12 weeks

[after_1980(0);highearn(1)]: pre-treatement treated group mean_log_duration : 1.38 weeks

[after_1980(1);highearn(0)]: post-treatement control group mean_log_duration : 1.13 weeks

[after_1980(1);highearn(1)]: post-treatement treated group mean_log_duration : 1.58 weeks

policy_effect =[(log_avg_duration(post-treatement treated group)-log_avg_duration(pre-treatement treated group)]-[log_avg_duration(post-treatement control group)-log_avg_duration(pre-treatement control group)]

policy_effect = [log_avg_duration of treated group (POST-PRE)]-[log_avg_duration of control group(POST-PRE)]

```r
pre_treatement_treated_group_log <- difr_log %>%
  filter(after_1980 == 0, highearn == 1) %>%
  pull(mean.log_duration)

pre_treatement_control_group_log<- difr_log %>%
  filter(after_1980 == 0, highearn == 0) %>%
  pull(mean.log_duration)

post_treatement_treated_group_log <- difr_log %>%
  filter(after_1980 == 1, highearn == 1) %>%
  pull(mean.log_duration)

post_treatement_control_group_log <- difr_log %>%
  filter(after_1980 == 1, highearn == 0) %>%
  pull(mean.log_duration)
```

```r
log_treatement_group_before_after <- post_treatement_treated_group_log -
  pre_treatement_treated_group_log
log_control_group_before_after    <-post_treatement_control_group_log-
  pre_treatement_control_group_log
```

Policy Effect log(DiD log Estimate:)

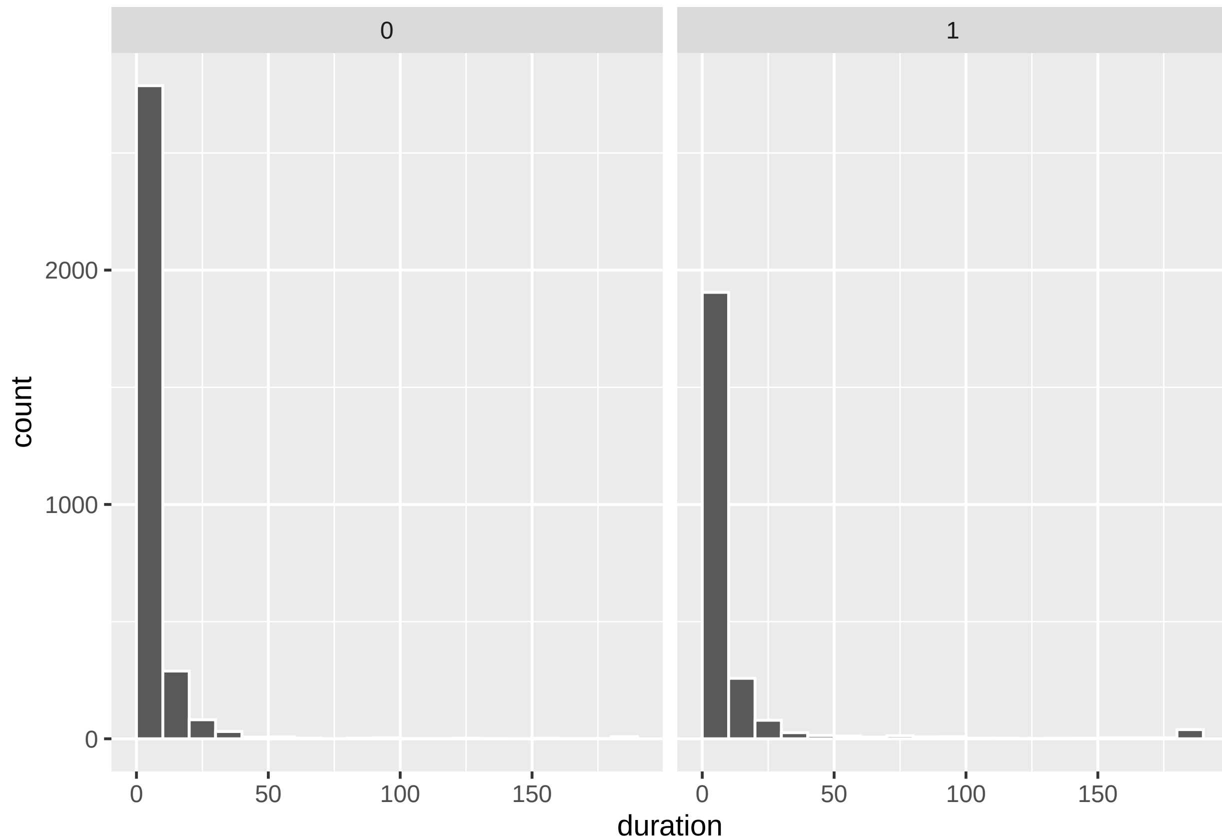
```
policy_effect_log <- log_treatement_group_before_after - log_control_group_before_after
print(policy_effect_log)
```

```
## [1] 0.1906012
```

Plotting

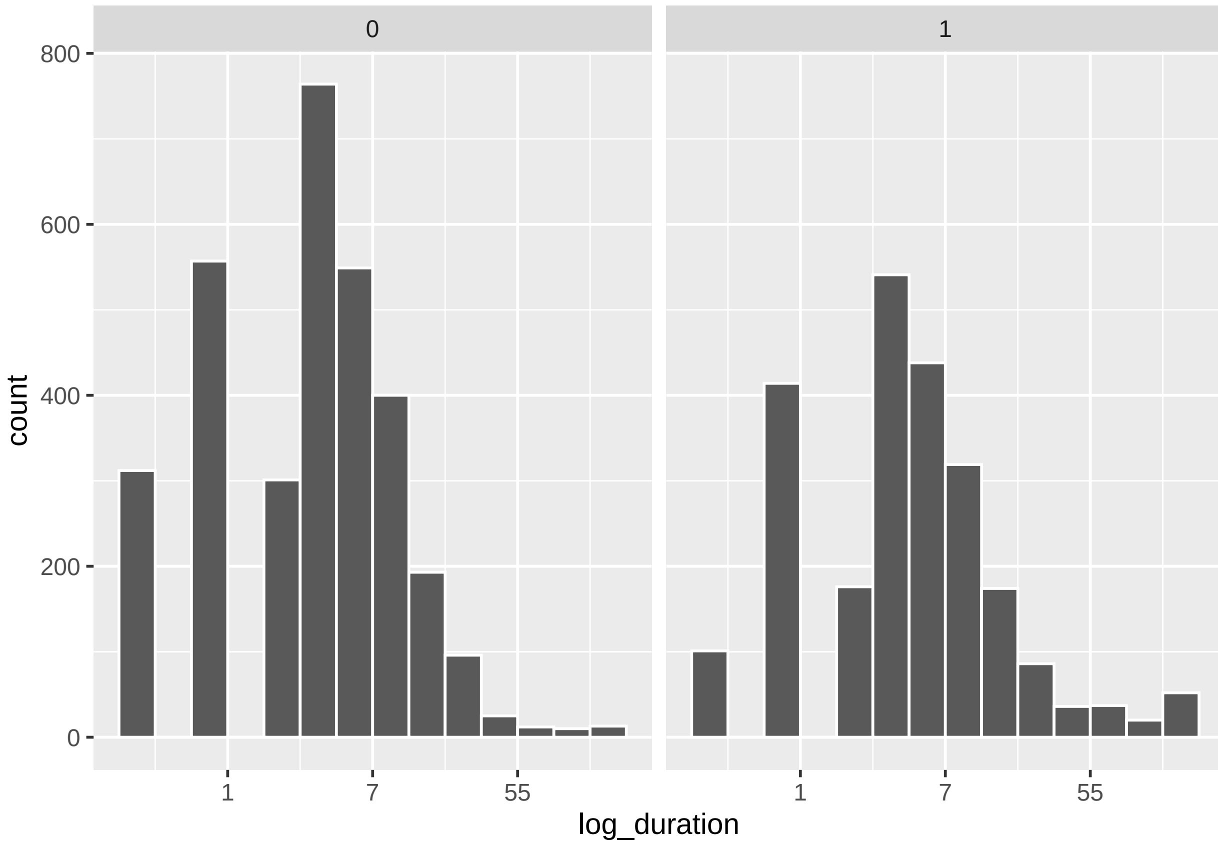
Distribution of Duration by Category

```
ggplot(data = df, aes(x = duration)) +
  geom_histogram(binwidth = 10, color = "white", boundary = 0) +
  facet_wrap(~ highearn)
```



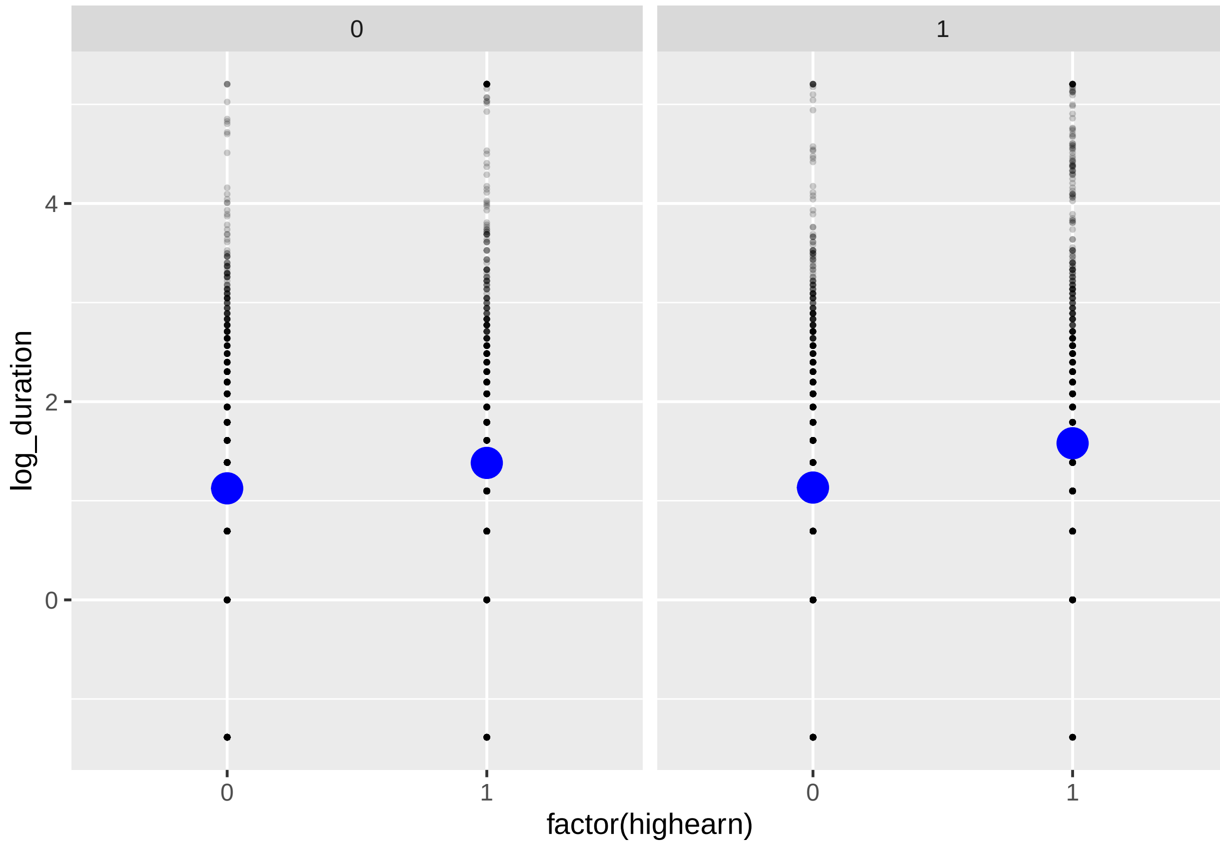
Ploting log duration based on category

```
ggplot(data = df, mapping = aes(x = log_duration)) +
  geom_histogram(binwidth = 0.5, color = "white", boundary = 0) +
   scale_x_continuous(labels = trans_format("exp", format = round)) +
  facet_wrap(~ highearn)
```



Here we just calculate Mean

```
ggplot(df, aes(x = factor(highearn), y = log_duration)) +
  geom_point(size = 0.5, alpha = 0.2) +
  stat_summary(geom = "point", fun = "mean", size = 5, color = "blue") +
  facet_wrap(vars(after_1980))
```



3.Basic Regression analysis to calculate the estimates without any control

Basic Regression model

$$duration = \beta_0 + \beta_1 after1980 + \beta_2 highearn + \delta afterchange \cdot highearn + u$$

```
basic_model_1 <- lm(duration ~ after_1980 + highearn + after_1980*highearn, data = df)
tidy(basic_model_1)
```

```
## # A tibble: 4 x 5
##   term                estimate std.error statistic  p.value
##   <chr>                  <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)            6.27      0.523    12.0   9.51e-33
## 2 after_1980             0.766     0.761     1.01  3.14e- 1
## 3 highearn               4.91      0.807     6.08  1.30e- 9
## 4 after_1980:highearn    0.951     1.17      0.816 4.14e- 1
```

The notation $\delta 1 = 0.9513$, indicates that the policy change might have increased the duration of benefits for high-income workers by about 0.9513 weeks more than for low-income workers. However, this effect is not statistically significant (p=0.414). The coefficient on after_1980 is small 0.7658 and statistically insignificant which means the increase in the earnings cap has no effect on duration for low-income workers.

$$\log(duration) = \beta_0 + \beta_1 after1980 + \beta_2 highearn + \delta_1 afterchange \cdot highearn + u$$

```r
basic_model_2 <- lm(log(duration) ~ after_1980 + highearn + after_1980*highearn, data = df)
tidy(basic_model_2)
```

```
## # A tibble: 4 x 5
##   term                estimate std.error statistic   p.value
##   <chr>                  <dbl>     <dbl>     <dbl>     <dbl>
## 1 (Intercept)          1.13      0.0307    36.6   1.62e-263
## 2 after_1980           0.00766   0.0447     0.171 8.64e-  1
## 3 highearn             0.256     0.0474     5.41  6.72e-  8
## 4 after_1980:highearn  0.191     0.0685     2.78  5.42e-  3
```

The notation δ signifies that the average duration of workers' compensation among high earners rose approximately by 19.06% due to the increased earnings cap. The coefficient on after_1980 is small 0.007 and statistically insignificant which means the increase in the earnings cap has no effect on duration for low-income workers.

4.regression adujstment procedure to evaluate the impact of policy change on the duration, and log_duration.

Regression Adjustment Model 1

$$\text{duration} = \beta_0 + \beta_1 \text{after1980} + \beta_2 \text{highearn} + \delta_1 \text{afterchange} \cdot \text{highearn}$$
$$+\gamma_1 \text{male} + \gamma_2 \text{married} + \gamma_3 \text{age}^2 + \gamma_4 \text{hosp} + \gamma_5 \text{indust}$$
$$+\gamma_6 \text{injtype} + \gamma_7 \text{lprewage} + u$$

Creating a new column age squared

```r
df <- df %>% mutate(age_squared = age^2)
```

```r
adv_model_1 <- lm(duration ~ after_1980 + highearn +
after_1980*highearn + male + married + age_squared +
hosp + indust + injtype + lprewage, data =df)
print(summary(adv_model_1))
```

```
##
## Call:
## lm(formula = duration ~ after_1980 + highearn + after_1980 *
##     highearn + male + married + age_squared + hosp + indust +
##     injtype + lprewage, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -27.603  -6.471  -2.251   1.503 181.142
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -1.863e+01  7.171e+00  -2.598 0.009390 **
## after_1980           1.101e+00  7.031e-01   1.566 0.117441
## highearn            -1.100e+00  1.517e+00  -0.726 0.468082
## male                -9.995e-01  7.165e-01  -1.395 0.163113
```

```
## married              1.581e+00  6.286e-01   2.515 0.011938 *
## age_squared          3.412e-04  2.820e-04   1.210 0.226412
## hosp                 1.346e+01  6.296e-01  21.379  < 2e-16 ***
## indust2              7.637e-01  9.210e-01   0.829 0.407036
## indust3              2.781e+00  6.440e-01   4.318  1.6e-05 ***
## injtype2             5.904e+00  2.446e+00   2.414 0.015813 *
## injtype3             1.104e+00  1.454e+00   0.760 0.447436
## injtype4            -5.895e-01  1.580e+00  -0.373 0.709034
## injtype5             3.461e+00  1.454e+00   2.381 0.017308 *
## injtype6             3.725e-01  1.469e+00   0.254 0.799757
## injtype7             1.071e+01  3.241e+00   3.303 0.000962 ***
## injtype8            -1.464e-02  2.023e+00  -0.007 0.994227
## lprewage             3.454e+00  1.362e+00   2.536 0.011226 *
## after_1980:highearn  1.790e+00  1.088e+00   1.644 0.100192
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.56 on 5329 degrees of freedom
##   (279 observations deleted due to missingness)
## Multiple R-squared:  0.1102, Adjusted R-squared:  0.1074
## F-statistic: 38.83 on 17 and 5329 DF,  p-value: < 2.2e-16
```

After controlling for all the X's (male ,married , age_squared ,hosp,indust,injtype,lprewage), the coeffcicent of interation term comes out to be 1.79 whic is insignificant.

Regression Adjustment Model 2

$$\log(\text{duration}) = \beta_0 + \beta_1\text{after1980} + \beta_2\text{highearn} + \delta_1\text{afterchange}\cdot\text{highearn}$$
$$+\gamma_1\text{male} + \gamma_2\text{married} + \gamma_3\text{age}^2 + \gamma_4\text{hosp} + \gamma_5\text{indust}$$
$$+\gamma_6\text{injtype} + \gamma_7\text{lprewage} + u$$

After controlling for all the X's (male ,married , age_squared ,hosp,indust, injtype,lprewage), the coeffcicent of interation term comes out to be 1.69 whic is significant , which means that the he average duration of workers' compensation among high earners rose approximately by 16.07% due to the increased earnings cap.

```r
adv_model_2 <- lm(log_duration ~ after_1980 + highearn +
after_1980*highearn + male + married + age_squared + hosp
+ indust + injtype + lprewage, data =df)
print(summary(adv_model_2))
```

```
##
## Call:
## lm(formula = log_duration ~ after_1980 + highearn + after_1980 *
##     highearn + male + married + age_squared + hosp + indust +
##     injtype + lprewage, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.0550 -0.7754  0.0930  0.7318  4.3620
##
## Coefficients:
```

```
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -1.460e+00  4.216e-01  -3.463 0.000538 ***
## after_1980            4.937e-02  4.134e-02   1.194 0.232437
## highearn             -1.532e-01  8.916e-02  -1.718 0.085832 .
## male                 -9.264e-02  4.213e-02  -2.199 0.027931 *
## married               6.771e-02  3.696e-02   1.832 0.067005 .
## age_squared           7.357e-05  1.658e-05   4.437 9.32e-06 ***
## hosp                  1.131e+00  3.702e-02  30.546  < 2e-16 ***
## indust2               1.831e-01  5.415e-02   3.380 0.000729 ***
## indust3               1.629e-01  3.787e-02   4.301 1.73e-05 ***
## injtype2              9.364e-01  1.438e-01   6.512 8.09e-11 ***
## injtype3              6.362e-01  8.547e-02   7.443 1.14e-13 ***
## injtype4              5.571e-01  9.288e-02   5.998 2.13e-09 ***
## injtype5              6.446e-01  8.547e-02   7.542 5.42e-14 ***
## injtype6              6.163e-01  8.634e-02   7.137 1.08e-12 ***
## injtype7              9.966e-01  1.906e-01   5.229 1.77e-07 ***
## injtype8              4.361e-01  1.190e-01   3.666 0.000248 ***
## lprewage              2.943e-01  8.005e-02   3.676 0.000239 ***
## after_1980:highearn   1.697e-01  6.400e-02   2.651 0.008043 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.15 on 5329 degrees of freedom
##   (279 observations deleted due to missingness)
## Multiple R-squared:  0.1893, Adjusted R-squared:  0.1867
## F-statistic: 73.17 on 17 and 5329 DF,  p-value: < 2.2e-16
```

Combining the data set to contain (ky==0) We here combine the data set including the state of michigan state sample .

```
dataset_combined <- dataset %>% rename(duration = durat,
log_duration = ldurat,after_1980 = afchnge)%>%
mutate(indust = as.factor(indust),injtype = as.factor(injtype))%>%
mutate(age_squared = age^2)
```

5. Robustness check

$$\log(\text{duration}) = \beta_0 + \beta_1\text{after1980} + \beta_2\text{highearn} + \delta_1\text{afterchange}\cdot\text{highearn}$$
$$+\gamma_1\text{male} + \gamma_2\text{married} + \gamma_3\text{age}^2 + \gamma_4\text{hosp} + \gamma_5\text{indust}$$
$$+\gamma_6\text{injtype} + \gamma_7\text{lprewage} + \delta_2\text{after1980}\cdot\text{ky} + \delta_3\text{ky} + u$$

```
robust_model1 <- lm(log_duration ~ after_1980 + highearn +
after_1980*highearn + male + married + age_squared + hosp
+ indust + injtype + lprewage + ky +ky:after_1980, data =dataset_combined)
print(summary(robust_model1))
```

```
##
## Call:
## lm(formula = log_duration ~ after_1980 + highearn + after_1980 *
##     highearn + male + married + age_squared + hosp + indust +
```

```
##     injtype + lprewage + ky + ky:after_1980, data = dataset_combined)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.7673 -0.7489  0.0699  0.7357  4.2709
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -1.175e+00  4.514e-01  -2.603 0.009270 **
## after_1980           8.059e-02  6.422e-02   1.255 0.209597
## highearn            -1.326e-01  8.690e-02  -1.526 0.126941
## male                -1.438e-01  3.852e-02  -3.732 0.000192 ***
## married              4.594e-02  3.324e-02   1.382 0.167018
## age_squared          8.785e-05  1.501e-05   5.853 5.04e-09 ***
## hosp                 1.101e+00  3.346e-02  32.910  < 2e-16 ***
## indust2              2.640e-01  4.724e-02   5.588 2.38e-08 ***
## indust3              1.551e-01  3.370e-02   4.603 4.24e-06 ***
## injtype2             8.898e-01  1.333e-01   6.674 2.68e-11 ***
## injtype3             6.550e-01  7.936e-02   8.254  < 2e-16 ***
## injtype4             5.969e-01  8.558e-02   6.975 3.35e-12 ***
## injtype5             6.315e-01  7.956e-02   7.938 2.39e-15 ***
## injtype6             6.068e-01  8.025e-02   7.562 4.49e-14 ***
## injtype7             1.115e+00  1.612e-01   6.918 5.00e-12 ***
## injtype8             5.605e-01  1.080e-01   5.190 2.16e-07 ***
## lprewage             2.758e-01  7.919e-02   3.483 0.000499 ***
## ky                  -1.628e-01  5.932e-02  -2.745 0.006064 **
## after_1980:highearn  1.674e-01  5.887e-02   2.843 0.004476 **
## after_1980:ky       -2.950e-02  6.997e-02  -0.422 0.673382
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.179 on 6802 degrees of freedom
##   (328 observations deleted due to missingness)
## Multiple R-squared:  0.1832, Adjusted R-squared:  0.1809
## F-statistic: 80.27 on 19 and 6802 DF,  p-value: < 2.2e-16
```

DDD

$$\log(\text{duration}) = \beta_0 + \beta_1 * ky + \beta_2 * after1980_t + \beta_3 * highEarn_i +$$

$$\gamma_1(ky_s * after1980_t) + \gamma_2(ky * highearn) + \gamma_2(after1980 * highearn) + \delta_1(ky * after1980 * highearn) + \psi * X + u$$

Considering the entire dataset including Michigan, we need to account for state-specific and income-specific time trends that might influence compensation benefits. To do this, we use Triple Differences, which involves comparing the differences between Kentucky and Michigan data to eliminate these confounding factors.

```
DDD= lm( duration ~ ky + after_1980 + highearn + ky*after_1980 + ky*highearn
         + after_1980*highearn + ky*after_1980*highearn + male + married
         + age_squared + hosp
+ indust + injtype + lprewage, data=dataset_combined)
summary(DDD)
```

```
##
```

```
## Call:
## lm(formula = duration ~ ky + after_1980 + highearn + ky * after_1980 +
##     ky * highearn + after_1980 * highearn + ky * after_1980 *
##     highearn + male + married + age_squared + hosp + indust +
##     injtype + lprewage, data = dataset_combined)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -41.970  -7.650  -2.916   1.287 179.648
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -1.483e+01  8.685e+00  -1.708 0.087693 .
## ky                      -3.296e+00  1.295e+00  -2.545 0.010941 *
## after_1980               2.708e+00  1.421e+00   1.906 0.056712 .
## highearn                 5.330e-01  2.303e+00   0.231 0.816988
## male                    -1.865e+00  7.407e-01  -2.518 0.011813 *
## married                  8.539e-01  6.387e-01   1.337 0.181296
## age_squared              8.690e-04  2.888e-04   3.009 0.002634 **
## hosp                     1.380e+01  6.430e-01  21.468  < 2e-16 ***
## indust2                  2.311e+00  9.102e-01   2.539 0.011141 *
## indust3                  2.485e+00  6.476e-01   3.838 0.000125 ***
## injtype2                 6.412e+00  2.562e+00   2.503 0.012331 *
## injtype3                 1.889e+00  1.525e+00   1.239 0.215275
## injtype4                 5.805e-01  1.644e+00   0.353 0.723992
## injtype5                 4.376e+00  1.528e+00   2.863 0.004204 **
## injtype6                 8.398e-01  1.542e+00   0.545 0.585941
## injtype7                 1.297e+01  3.101e+00   4.181 2.94e-05 ***
## injtype8                 2.665e+00  2.075e+00   1.284 0.199142
## lprewage                 3.250e+00  1.522e+00   2.136 0.032710 *
## ky:after_1980           -1.636e+00  1.638e+00  -0.999 0.317866
## ky:highearn             -1.269e+00  1.975e+00  -0.643 0.520521
## after_1980:highearn      5.057e-01  2.576e+00   0.196 0.844378
## ky:after_1980:highearn   1.361e+00  2.868e+00   0.474 0.635199
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22.64 on 6800 degrees of freedom
##   (328 observations deleted due to missingness)
## Multiple R-squared:  0.09577,    Adjusted R-squared:  0.09298
## F-statistic:  34.3 on 21 and 6800 DF,  p-value: < 2.2e-16
```

here we see that the triple interaction term comes out to be 1.36 positive but insignificant, when controlling for all the other possible X's.

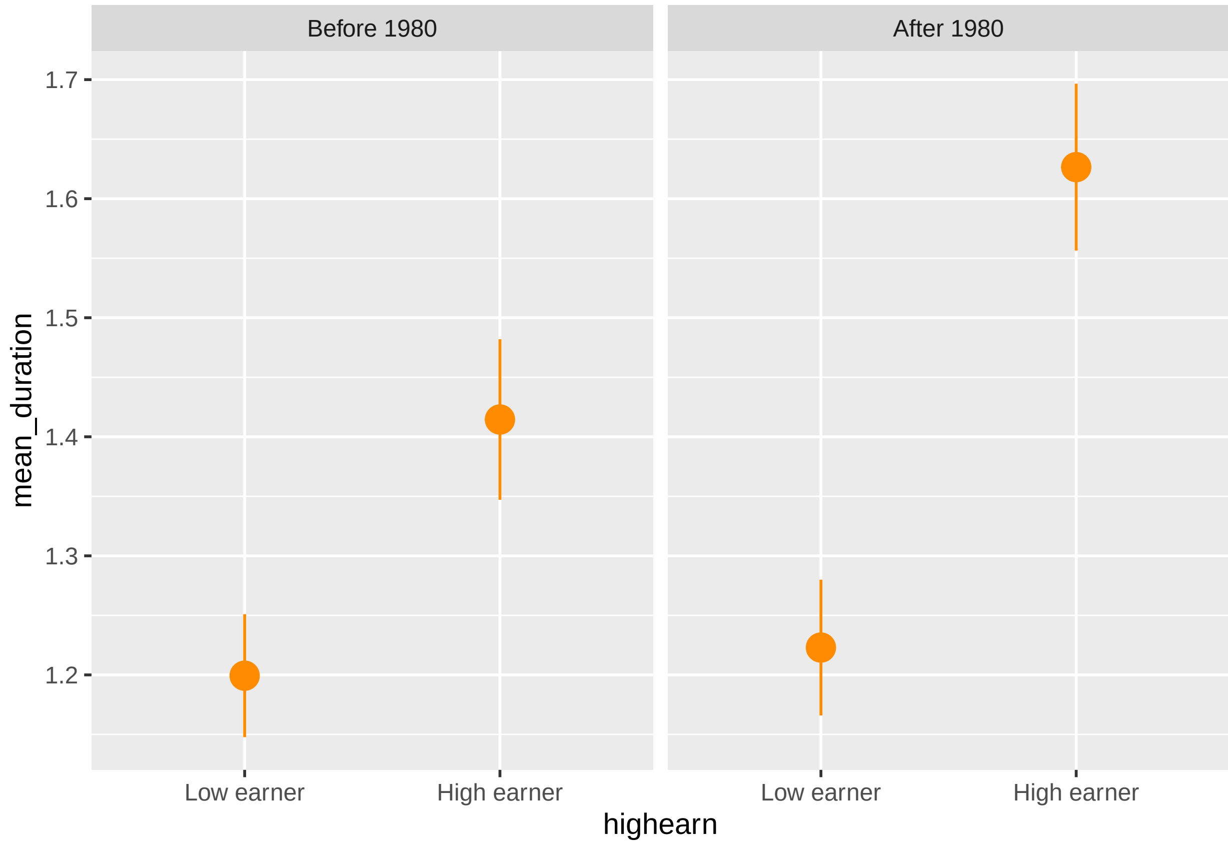
6. Generating Graph

```
graph_data <- dataset_combined %>%
   mutate(highearn = factor(highearn, labels = c("Low earner", "High earner")),
         after_1980 = factor(after_1980, labels = c("Before 1980", "After 1980"))) %>%
  group_by(highearn, after_1980) %>%
```

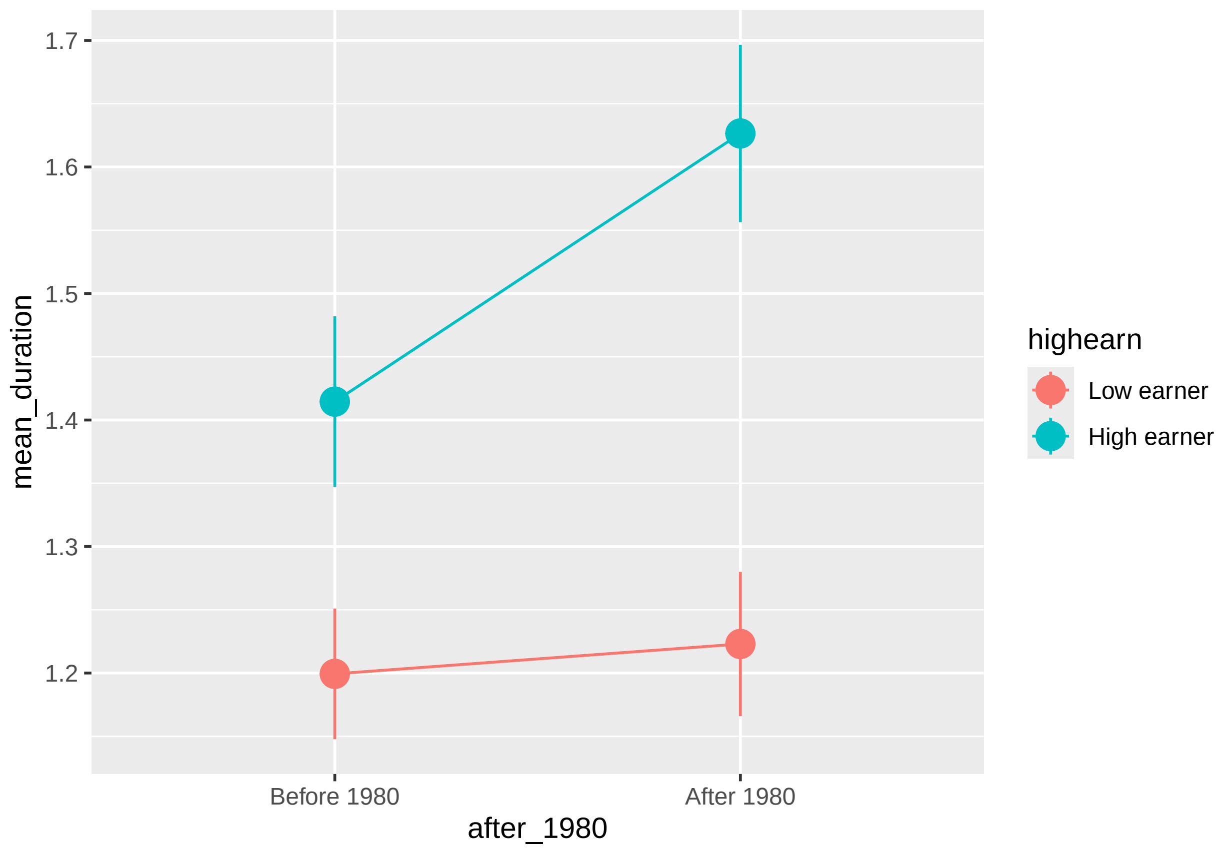
```
  summarize(mean_duration = mean(log_duration),
            se_duration = sd(log_duration) / sqrt(n()),
            upper = mean_duration + (1.96 * se_duration),
            lower = mean_duration + (-1.96 * se_duration))
```

```
## `summarise()` has grouped output by 'highearn'. You can override using the
## `.groups` argument.
```

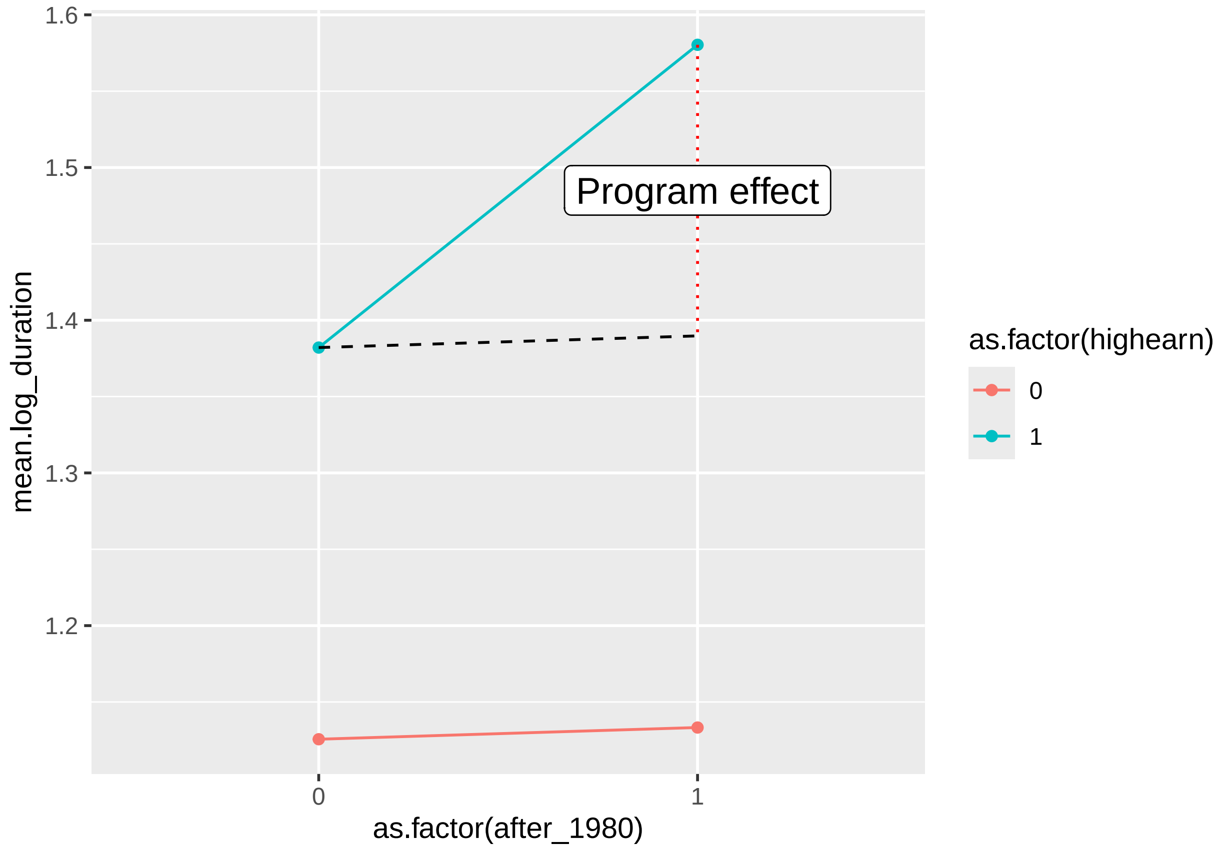
```
ggplot(graph_data, aes(x = highearn, y = mean_duration)) +
  geom_pointrange(aes(ymin = lower, ymax = upper),
                  color = "darkorange", size = 1) +
  facet_wrap(vars(after_1980))
```




```
ggplot(graph_data, aes(x = after_1980, y = mean_duration, color = highearn)) +
  geom_pointrange(aes(ymin = lower, ymax = upper), size = 1) +
  # The group = highearn here makes it so the lines go across categories
  geom_line(aes(group = highearn))
```




```r
ggplot(difr_log, aes(x = as.factor(after_1980),
                     y = mean.log_duration,
                     color = as.factor(highearn)))+
  geom_point() +
  geom_line(aes(group = as.factor(highearn))) +
  annotate(geom = "segment", x = "0", xend = "1",
           y = pre_treatement_treated_group_log, yend =
             post_treatement_treated_group_log - policy_effect_log,
           linetype = "dashed", color = "black") +
  annotate(geom = "segment", x = "1", xend = "1",
           y = post_treatement_treated_group_log, yend =
             post_treatement_treated_group_log - policy_effect_log,
           linetype = "dotted", color = "red") +
  annotate(geom = "label", x = "1", y = post_treatement_treated_group_log
           - (policy_effect_log / 2),
           label = "Program effect", size = 5)
```




```
modelsummary(list(basic_model_2,adv_model_2,DDD,robust_model1),output='latex')
```

	(1)	(2)	(3)	(4)
(Intercept)	1.126	−1.460	−14.834	−1.175
	(0.031)	(0.422)	(8.685)	(0.451)
after_1980	0.008	0.049	2.708	0.081
	(0.045)	(0.041)	(1.421)	(0.064)
highearn	0.256	−0.153	0.533	−0.133
	(0.047)	(0.089)	(2.303)	(0.087)
after_1980 × highearn	0.191	0.170	0.506	0.167
	(0.069)	(0.064)	(2.576)	(0.059)
male		−0.093	−1.865	−0.144
		(0.042)	(0.741)	(0.039)
married		0.068	0.854	0.046
		(0.037)	(0.639)	(0.033)
age_squared		0.000	0.001	0.000
		(0.000)	(0.000)	(0.000)
hosp		1.131	13.804	1.101
		(0.037)	(0.643)	(0.033)
indust2		0.183	2.311	0.264
		(0.054)	(0.910)	(0.047)
indust3		0.163	2.485	0.155
		(0.038)	(0.648)	(0.034)
injtype2		0.936	6.412	0.890
		(0.144)	(2.562)	(0.133)
injtype3		0.636	1.889	0.655
		(0.085)	(1.525)	(0.079)
injtype4		0.557	0.581	0.597
		(0.093)	(1.644)	(0.086)
injtype5		0.645	4.376	0.631
		(0.085)	(1.528)	(0.080)
injtype6		0.616	0.840	0.607
		(0.086)	(1.542)	(0.080)
injtype7		0.997	12.966	1.115
		(0.191)	(3.101)	(0.161)
injtype8		0.436	2.665	0.561
		(0.119)	(2.075)	(0.108)
lprewage		0.294	3.250	0.276
		(0.080)	(1.522)	(0.079)
ky			−3.296	−0.163
			(1.295)	(0.059)
ky × after_1980			−1.636	
			(1.638)	
ky × highearn			−1.269	
			(1.975)	
ky × after_1980 × highearn			1.361	
			(2.868)	