IO & Econometrics workshop - PS1

Doron Zamir

This is a markdown file created using RStudio.

Using Packeges

I used these packages to create the data:

```
library(tidyverse)
library(knitr)
library(magrittr)
library(AER)
library(broom)
```

Tidy the Data

Importing Data

I used the data from David Card's website (https://davidcard.berkeley.edu/data_sets.html) The relevant file is `nls.dat` saved in the `Data` directory.

```
data <- read.table("Data/nls.dat") %>% as_data_frame()
head(data)
```

V1 <int>	V2 <int>	V3 <int>	V4 <int>	V5 <int>	V6 <int>	V7 <int>	V8 <int>	V9 <dbl>	V10 <int>
2	0	0	0	0	7	5	29	9.94	1
3	0	0	0	0	12	11	27	8.00	0
4	0	0	0	0	12	12	34	14.00	0
5	1	1	1	0	11	11	27	11.00	0
6	1	1	1	0	12	12	34	8.00	0
7	1	1	1	0	12	11	26	9.00	0

6 rows | 1-10 of 52 columns

Changing col names & type

Using variable names from `code_bk.txt'

```
colnames(data) <-(c("id","nearc2","nearc4","nearc4a","nearc4b","ed76","ed66","age7
6","daded","nodaded","momed","nomomed","weight","momdad14","sinmom14","step14","reg66
1","reg662","reg663","reg664","reg665","reg666","reg667","reg668","reg669","south6
6","work76","work78","lwage76","lwage78","famed","black","smsa76r","smsa78r","reg76
r","reg78r","reg80r","smsa66r","wage76","wage78","wage80","noint78","noint80","enroll
76","enroll78","enroll80","kww","iq","marsta76","marsta78","marsta80","libcrd14"))

data %<>% mutate_if(is_character,suppressWarnings(as.numeric))
head(data)
```

id <int>	nearc2 <int>	nearc4 <int>	nearc4a <int>	nearc4b <int>	ed76 <int>	ed66 <int>	age76 <int>	daded <dbl>	nodaded <int>
2	0	0	0	0	7	5	29	9.94	1
3	0	0	0	0	12	11	27	8.00	0
4	0	0	0	0	12	12	34	14.00	0
5	1	1	1	0	11	11	27	11.00	0
6	1	1	1	0	12	12	34	8.00	0
7	1	1	1	0	12	11	26	9.00	0

6 rows | 1-10 of 52 columns

Creating new Varibales for experiance

$$exp76 = age76 - ed76 - 6$$

$$exp762 = exp76 * exp76$$

```
data <- data %>% mutate(exp76 =  age76-ed76 - 6)
data <- data %>% mutate(exp762 = exp76 ** 2)
```

Sumaarise Data

```
data_summary <- data[-1] %>%
  summarise_all(list(
    Min = min,
    Mean = mean,
    Max = max,
    SD = sd)) %>%
  pivot_longer(everything(),
               names_to = c("Var","Stat"),
               names_sep = "_") %>%
  pivot_wider(names_from = "Stat") %>% column_to_rownames("Var")
data_summary %>% format(scientific = FALSE, digits = 2,trim = TRUE)
```

	Min <l<chr>>	Mean <l<chr>>	Max <l<chr>>	SD <l<chr>>
nearc2	0	0.432	1	0.50
nearc4	0	0.678	1	0.47
nearc4a	0	0.492	1	0.50
nearc4b	0	0.186	1	0.39
ed76	0	13.225	18	2.75
ed66	0	10.743	18	2.46
age76	24	28.175	34	3.17
daded	0	10.003	18	3.30
nodaded	0	0.224	1	0.42
momed	0	10.342	18	3.03

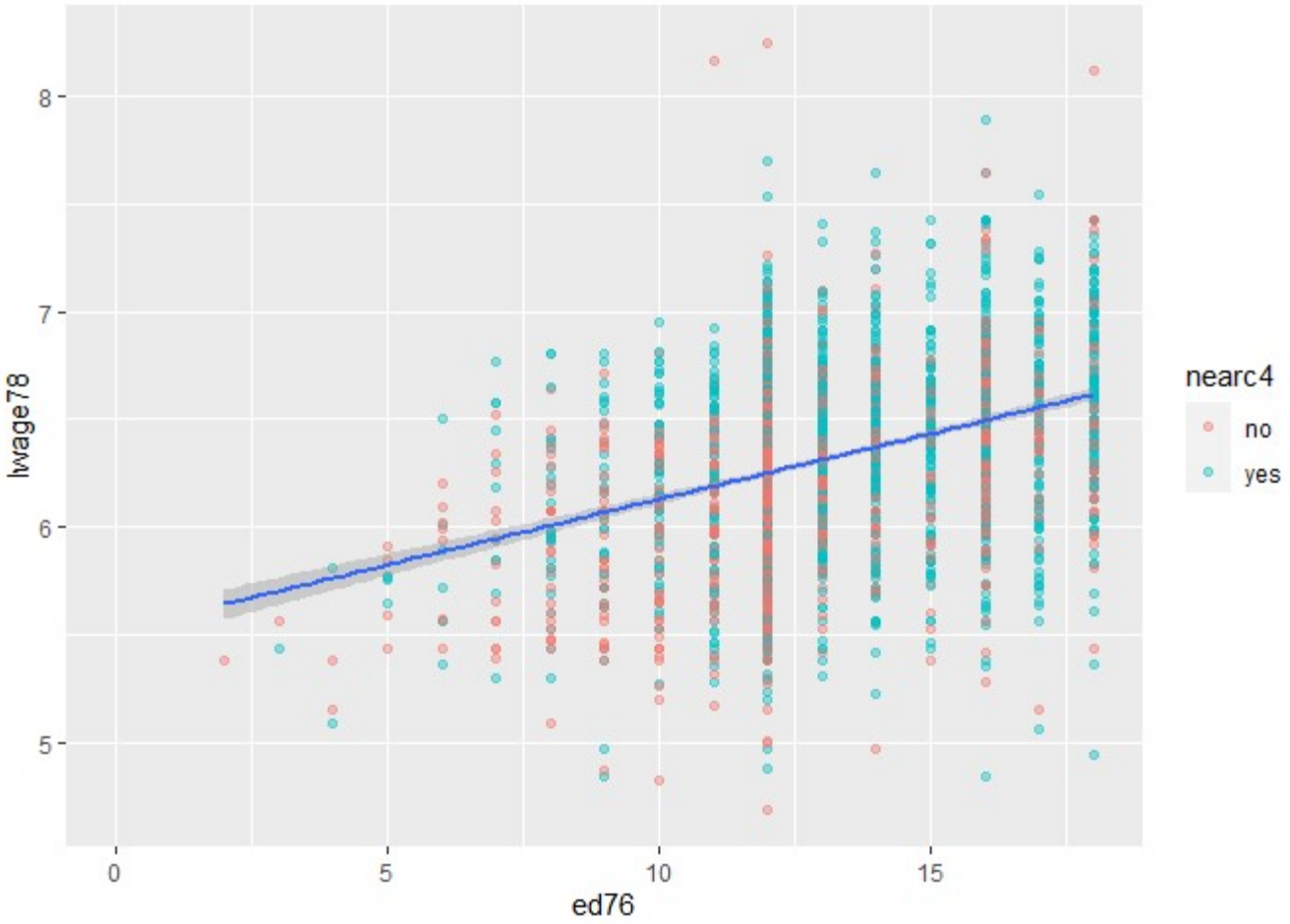
1-10 of 53 rows Previous **1** 2 3 4 5 6 Next

Plotting

```
data_to_plot <- data %>%
  mutate(
    black = factor(black,labels = c("no","yes")),
    nearc4=factor(nearc4,labels = c("no","yes")),
         )
```

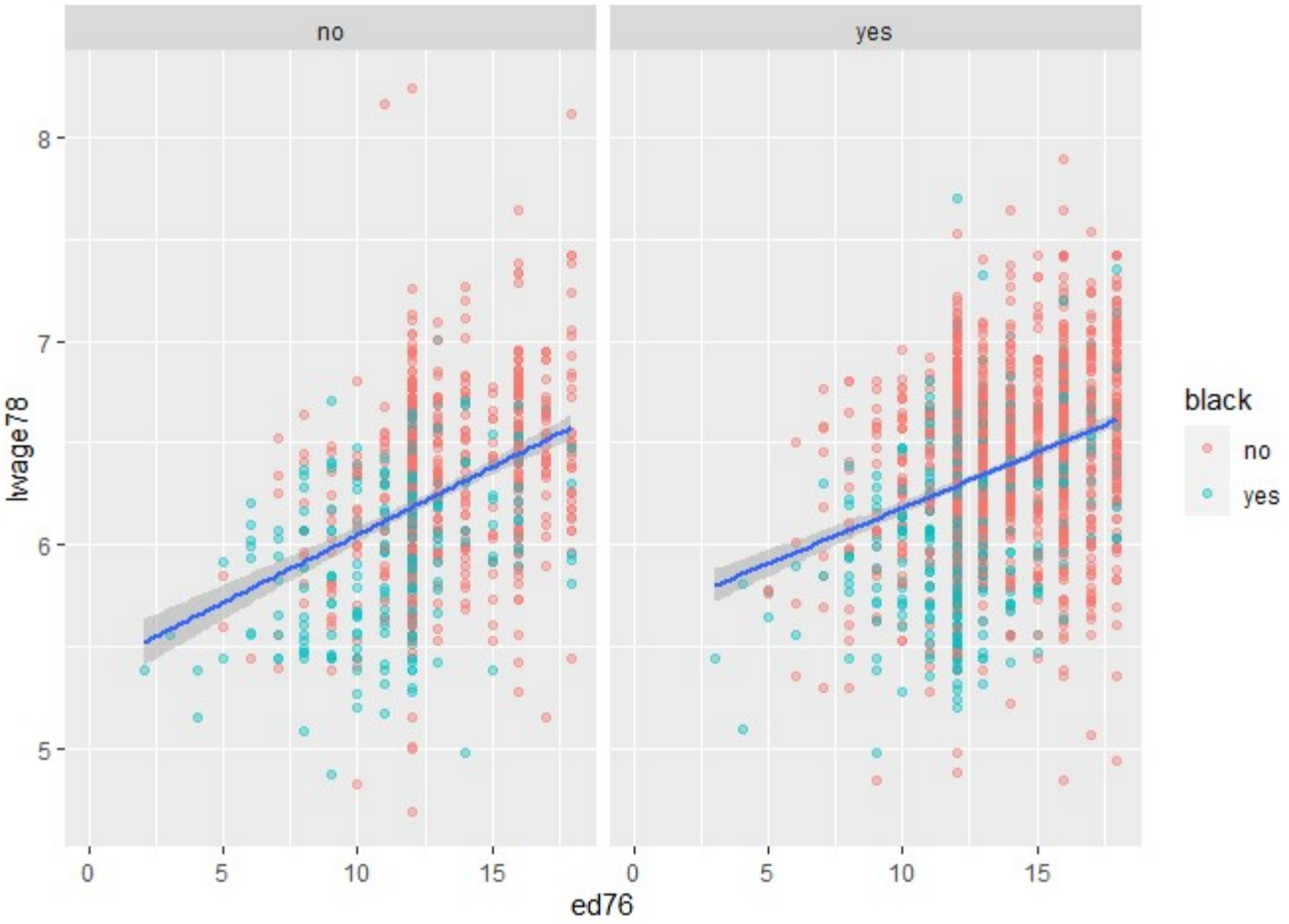
Education and lwage

```
data_to_plot %>% ggplot(aes(ed76,lwage78)) +
  geom_point(aes(
    color = nearc4),
    alpha =0.4 ) +
  geom_smooth(method = lm)
```



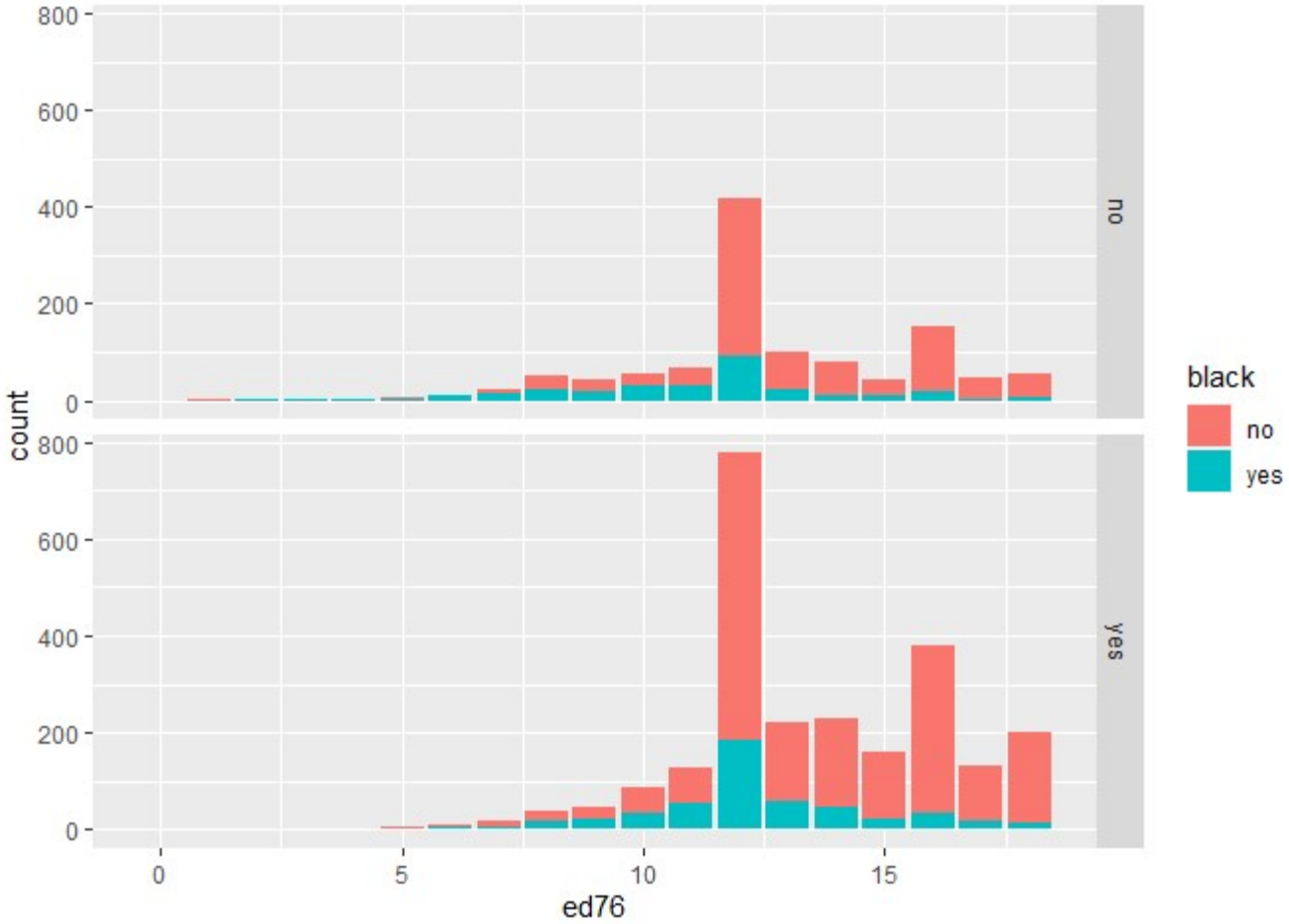
Splitting by distance from collage

```
data_to_plot %>% ggplot(aes(ed76,lwage78)) +
  geom_point(aes(
    color = black),
    alpha =0.4 ) +
  geom_smooth(method = lm)+
  facet_grid(cols = vars(nearc4))
```



Histogram of ed76

```
data_to_plot %>% ggplot(aes(ed76)) +
  geom_bar(aes(
    fill = black)) +
  facet_grid(row = vars(nearc4))
```



Estimating a model

OLS

$$lwage78 = \alpha \cdot ed76 + \beta \cdot X + u$$

I used model #2 from Card's paper

```
ols_model <- data %>% lm(lwage78~ed76 + exp76 +exp762+smsa76r+ reg76r+smsa66r+
+reg662+reg663+reg664+reg665+reg666+reg667+reg668+reg669, data = .)
kable(tidy(ols_model))
```

term	estimate	std.error	statistic	p.value
(Intercept)	4.7555341	0.0824573	57.6726702	0.0000000
ed76	0.0778239	0.0038744	20.0865331	0.0000000
exp76	0.0710414	0.0073298	9.6920978	0.0000000
exp762	-0.0021642	0.0003494	-6.1947749	0.0000000
smsa76r	0.1493874	0.0222527	6.7132142	0.0000000
reg76r	-0.0977106	0.0291096	-3.3566420	0.0008002

term	estimate	std.error	statistic	p.value
smsa66r	0.0069889	0.0216880	0.3222488	0.7472899
reg662	0.0794742	0.0389442	2.0407198	0.0413786
reg663	0.1096911	0.0379474	2.8906086	0.0038766
reg664	0.0148938	0.0451717	0.3297151	0.7416416
reg665	-0.0118675	0.0448938	-0.2643466	0.7915336
reg666	0.0272414	0.0489813	0.5561604	0.5781486
reg667	0.0289232	0.0485522	0.5957131	0.5514183
reg668	-0.0069862	0.0550232	-0.1269678	0.8989756
reg669	0.1043800	0.0419997	2.4852578	0.0130076

2SLS

$$lwage78 = \alpha \cdot (\delta \cdot nearc4 + v) + \beta \cdot X + u$$

```
IV_model <- data %>% ivreg(lwage78~ed76 + exp76 +exp762+smsa76r+reg76r+smsa66r+reg66
2+reg663+reg664+reg665+reg666+reg667+reg668+reg669|exp76 +exp762+smsa76r+reg76r+smsa6
6r+reg662+reg663+reg664+reg665+reg666+reg667+reg668+reg669 + nearc4, data = .)
kable(tidy(IV_model))
```

term	estimate	std.error	statistic	p.value
(Intercept)	4.6448679	1.0660373	4.3571344	0.0000137
ed76	0.0843520	0.0628153	1.3428567	0.1794345
exp76	0.0738043	0.0275303	2.6808421	0.0073897
exp762	-0.0021724	0.0003583	-6.0632186	0.0000000
smsa76r	0.1465746	0.0350071	4.1869963	0.0000292
reg76r	-0.0980351	0.0292916	-3.3468673	0.0008288
smsa66r	0.0060301	0.0235729	0.2558053	0.7981212
reg662	0.0817170	0.0445227	1.8353996	0.0665596
reg663	0.1112795	0.0409182	2.7195607	0.0065799
reg664	0.0152388	0.0453174	0.3362676	0.7366960
reg665	-0.0064019	0.0690874	-0.0926636	0.9261779
reg666	0.0344006	0.0844349	0.4074217	0.6837315
reg667	0.0353821	0.0787891	0.4490729	0.6534161

term	estimate	std.error	statistic	p.value
reg668	-0.0097824	0.0612538	-0.1597031	0.8731272
reg669	0.1041498	0.0420806	2.4750089	0.0133860

I used a standard model of wage as function of education, controlling experince ($ exp^2) and using panel data for race, region of living, etc.

For IV, I used procimity to collage (as card did). It sounds like a reasnble IV as it is corelated with education, but it can fail. It can effect wage and earnings in more ways the just throw education: it can effect the quality of education, the enveirmont a perso grows in, etc.