Title

Subtitle

Your Name

March 15, 2020

Contents

1 Section 1 (level 1)

In this section, we...

1.1 Subsection (level 2)

In this subsection, we...

1.1.1 Sub-subsection (level 3)

In this sub-subsection, we...

The symbol # makes the section title. Depending on the number of #, you can create different levels of sections.

2 Background

2.1 History of Japan

In this subsection, we...

3 Citation

You can cite articles such as papers and reports using bibtex. If you save the reference called Greenstone2015, put this name after @. For example, @Greenstone2015 can produce the following: Greenstone and Jack (2015). If you use brackets [], it creates (Donaldson and Storeygard, 2016).

Finally, when you want to cite multiple articles, sepalate the names using ;. It produces the following (Greenstone and Jack, 2015; Donaldson and Storeygard, 2016).

Note that you have to specify the citation style in the preamble of the Rmd. See this files preamble.

4 .Rprofile

.Rprofile can be executed by `source` with `include=FALSE` without showing the results in the output (like PDF). In this file I put the file path just after the preamble. It is very useful to run multiple `library` thet you often use.

5 Regression

We conduct regression without specifying the types of the standard error.

5.1 Set Up Data

```
data(CASchools)

# student teacher ratio
CASchools$STR <- CASchools$students / CASchools$teachers

# average test score
CASchools$score <- (CASchools$read + CASchools$math)/2
```

This is same as: "reg score STR" in Stata. In R language we always need to specify which data will be used.

```
fit1 <- lm(score ~ STR, data = CASchools)
summary(fit1)
```

```
##
## Call:
## lm(formula = score ~ STR, data = CASchools)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -47.73 -14.25   0.48  12.82  48.54
##
## Coefficients:
##             Estimate Std. Error t value             Pr(>|t|)
## (Intercept)   698.93       9.47   73.82 < 0.0000000000000002 ***
## STR            -2.28       0.48   -4.75            0.0000028 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.6 on 418 degrees of freedom
## Multiple R-squared:  0.0512, Adjusted R-squared:  0.049
## F-statistic: 22.6 on 1 and 418 DF,  p-value: 0.00000278
```

Present the equations.

```r
equatiomatic::extract_eq(fit1)
```

$$\text{score} = \alpha + \beta_1(\text{STR}) + \epsilon$$

```r
equatiomatic::extract_eq(fit1, use_coefs = TRUE)
```

$$\text{score} = 698.93 - 2.28(\text{STR}) + \epsilon$$

We can include further controls in the equation.

```r
fit2 <- lm(score ~ STR + english + income, data = CASchools)
summary(fit2)
```

```
##
## Call:
## lm(formula = score ~ STR + english + income, data = CASchools)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -42.80  -6.86   0.27   6.59  31.20
##
## Coefficients:
##              Estimate Std. Error t value            Pr(>|t|)
## (Intercept) 640.3155     5.7749  110.88 <0.0000000000000002 ***
## STR          -0.0688     0.2769   -0.25                 0.8
## english      -0.4883     0.0293  -16.67 <0.0000000000000002 ***
## income        1.4945     0.0748   19.97 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.3 on 416 degrees of freedom
## Multiple R-squared:  0.707,  Adjusted R-squared:  0.705
## F-statistic:  335 on 3 and 416 DF,  p-value: <0.0000000000000002
```

```r
equatiomatic::extract_eq(fit2)
```

$$\text{score} = \alpha + \beta_1(\text{STR}) + \beta_2(\text{english}) + \beta_3(\text{income}) + \epsilon$$

```r
equatiomatic::extract_eq(fit2, use_coefs = TRUE)
```

$$\text{score} = 640.32 - 0.07(\text{STR}) - 0.49(\text{english}) + 1.49(\text{income}) + \epsilon$$

5.2 Cluster Robust Standard Error

In empirical work we always deal with correlation within a group by clustering SE. Above lm code conduct regression under the assumption of homoskedasticity just like reg and without robust option in Stata. We now use the estimatr package.

```r
fit3 <- estimatr::lm_robust(score ~ STR + english + income,
                            clusters = county, se_type = "stata",
                            data = CASchools)
summary(fit3)
```

```
##
## Call:
## estimatr::lm_robust(formula = score ~ STR + english + income,
##     data = CASchools, clusters = county, se_type = "stata")
##
## Standard error type:  stata
##
## Coefficients:
##             Estimate Std. Error t value
## (Intercept) 640.3155     6.2866 101.855
## STR          -0.0688     0.2898  -0.237
## english      -0.4883     0.0309 -15.779
## income        1.4945     0.1023  14.604
##                                                          Pr(>|t|)
## (Intercept) 0.000000000000000000000000000000000000000000000000000697
## STR         0.813527324296684395399381628521950915455818176269531250000
## english     0.0000000000000000009930987090849287295359433197705020740
## income      0.0000000000000000175194413389987415730081432545853544527
##             CI Lower CI Upper DF
## (Intercept)  627.646  652.985 44
## STR           -0.653    0.515 44
## english       -0.551   -0.426 44
## income         1.288    1.701 44
##
## Multiple R-squared:  0.707 , Adjusted R-squared:  0.705
## F-statistic:  301 on 3 and 44 DF,  p-value: <0.0000000000000002
```

We estimated the same model as fit2, but clustered SE at the county level (we do not know whether this is the best unit). Clustered SE for STR increased from 0.277 to 0.299.

5.3 (Automatically) Create Regression Tables

Like `outreg2` in Stata, there are several ways to create tables in R using latex. You can also make tables in the MS Word file if you speficy in the preamble. The most popular one is `stargazer`, but in some situations `texreg` and `Hmisc` are more convenient.

5.4 Produce the Regression Table (with stargazer)

```r
# Define removed unnecessary variables from the table

# Define the row names
label_regressors <- c("Student Teacher Ratio", "English Score", "Parents Income")

stargazer(fit1, fit2,
          se = starprep(fit1, fit2,
                        clusters = CASchools$county,
                        se_type = "stata"),
```

```
          title = "OLS estimation",
          dep.var.caption = "",
          dep.var.labels = "Dependent Variable Name Here",
          covariate.labels = label_regressors,
          omit.stat = c("adj.rsq", "f", "ser"),
          notes.append = FALSE,
          # single.row = TRUE,
          digits = 4,
          notes = "",
          notes.align = "c",
          # align = T,
          no.space = T,
          add.lines = list(
            c("Log of population", "", "Yes"),
            c("Temperature", "", "Yes"),
            c("Demographic characteristic", "Yes", "Yes"),
            c("Household wealth", "Yes", "Yes")),
          # table.layout ="-n",
          # font.size = "scriptsize"
          header = FALSE
          )
```

Table 1: OLS estimation

	Dependent Variable Name Here	
	(1)	(2)
Student Teacher Ratio	-2.2800^{***}	-0.0688
	(0.7527)	(0.2898)
English Score		-0.4883^{***}
		(0.0309)
Parents Income		1.4940^{***}
		(0.1023)
Constant	698.9000^{***}	640.3000^{***}
	(15.7600)	(6.2870)
Log of population		Yes
Temperature		Yes
Demographic characteristic	Yes	Yes
Household wealth	Yes	Yes
Observations	420	420
R^2	0.0512	0.7072

Note:

References

Donaldson, D. and Storeygard, A. (2016). The View from Above: Applications of Satellite Data in Economics. *Journal of Economic Perspectives*, 30(4):171–98.

Greenstone, M. and Jack, B. K. (2015). Envirodevonomics: A Research Agenda for an Emerging Field. *Journal of Economic Literature*, 53(1):5–42.