Review Session 2

FWL, R^2, Standard errors, and Dummy Variables

Ben Berger

2/3/23

Frisch-Waugh-Lovell Theorem

Suppose we are interested in the coeffcient estimate $\hat{\beta}_1$ from the following SRF:

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{u} \tag{1}$$

Instead of estimating that regression, suppose we estimate these two:

$$X_1 = \hat{\gamma}_0 + \hat{\gamma}_1 X_2 + \hat{v} \tag{2}$$
$$Y = \hat{\alpha}_0 + \hat{\alpha}_1 \hat{v} + \hat{w} \tag{3}$$

- First is a regression of X_1 (explanatory variable of interest) on X_2 (control variable).
- Second is a regression of the residuals of Y on the residuals of the first regression.

Frisch-Waugh-Lovell Theorem

FWL states that $\hat{\alpha}_1 = \hat{\beta}_1$. The coefficient from the bivariate regression of Y on $\hat{v}$ — the part of X_1 that is not explained by X_2 — is equal to the coefficient on X_1 in the multiple regression of Y on X_1 and X_2.

Intution: $\hat{v}$ comprise the variation in X_1 that is unexplained by the control variable. This is the same variation that is used when we hold X_2 fixed in the multiple regression. Thus it has the same association with Y and we estimate the same coefficient.

FWL Example: Miles per gallon

We can use the mtcars data to estimate the regression of a car's miles per gallon (mpg) on its weight (wt) and number of cylinders (cyl).

```
# Fit multivariate regression
fit_multivariate <- lm(mpg ~ wt + cyl, mtcars)
print(fit_multivariate)
```

```
Call:
lm(formula = mpg ~ wt + cyl, data = mtcars)

Coefficients:
(Intercept)           wt          cyl  
     39.686       -3.191       -1.508  
```

FWL Example: Miles per gallon

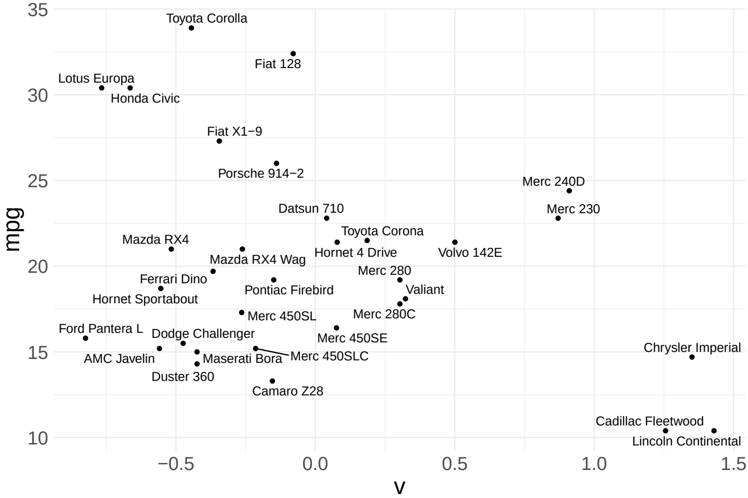
Fit the "auxillary" regression of `wt` on `cyl` and save the residuals

```
fit_auxillary <- lm(wt ~ cyl, mtcars)
print(fit_auxillary)
```

```

Call:
lm(formula = wt ~ cyl, data = mtcars)

Coefficients:
(Intercept)          cyl
     0.5646       0.4287
```

```
mtcars$v <- fit_auxillary$residuals
```

FWL Example: Miles per gallon



FWL Example: Miles per gallon

```r
fit_residuals <- lm(mpg ~ v, mtcars)
print(fit_residuals)
```

```
Call:
lm(formula = mpg ~ v, data = mtcars)

Coefficients:
(Intercept)            v
     20.091       -3.191
```

R^2

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

Share of the variation in the outcome that is explained (ESS) to the total variation from the mean (TSS). Equivalently, 1 minus the fraction that is unexplained (RSS/TSS).

R^2 close to 1 indicates that most of the variation is explained. If all points are on the regression line,[1] then R^2 is exactly 1.

R^2 close to zero indicates that little variation is explained.

What happens to R^2 when you add a covariate to a regression?

- It MUST increase. If you give the model more data, the worst it can do is explain the same amount of variation as before. Use **adjusted** R^2 to penalize adding covariates.

[1] ... or in multivariate regression with k covariates, the k-dimensional hyperplane

Statistical Inference for Regression Coefficients

- For a bivariate linear regression, $Y_i = \beta_0 + \beta_1 X_i + u_i$, we derived the OLS estimator for the slope coefficient $\hat{\beta}_1 = \frac{Cov(Y_i, X_i)}{Var(X_i)}$.
  - How precise is this?
  - Can we reject $\beta_1 = 0$?
- Use t-statistic like last semester: $t = \frac{\hat{\beta}_1 - 0}{se(\hat{\beta}_1)}$.
- What is $se(\hat{\beta}_1)$? Depends on the error structure of u_i.

Variance-covariance matrix

$$\begin{pmatrix} \sigma_1 & & & & \\ \sigma_{1,2} & \sigma_2 & & & \\ \sigma_{1,3} & \sigma_{2,3} & \sigma_3 & & \\ \vdots & \vdots & \vdots & \ddots & \\ \sigma_{1,N} & \sigma_{2,N} & \sigma_{3,N} & \dots & \sigma_N \end{pmatrix}$$

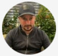
- Diagonal: person i's error variance.
  - Thought experiment: if I take an exam in 1000 parallel worlds, how much does the unexplained portion of my exam score vary from world to world? What about someone else's score?
  - Maybe I do very well if the day is sunny and very poorly if the day is rainy, while other people's performance is essentially unchanged.
- Off-diagonal: covariance of i's and j's errors.
  - Thought experiment: if my friend and I take an exam in 1000 parallel worlds, how much do the unexplained portions of our exam scores covary? What about my friend and someone who attends a different school?

Conventional standard errors

- Need to assume or estimate population (co)variances. Simplest assumption: conventional/IID/homoskedastic standard errors:

$$\begin{pmatrix} \sigma & & & & \\ 0 & \sigma & & & \\ 0 & 0 & \sigma & & \\ \vdots & \vdots & \vdots & \ddots & \\ 0 & 0 & 0 & \dots & \sigma \end{pmatrix}$$

- Assumes all errors are drawn from a distribution with variance of σ and zero correlation between individuals.
- $se_{conv}(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^N x_i^2}} = \sqrt{\frac{\frac{1}{N-1}\sum_{i=1}^N \hat{u}_i^2}{\sum_{i=1}^N x_i^2}}$
- Default standard errors reported by `lm()`, but in practice use one of the other types of errors.



Nate Silver @NateSilver538 · Jun 21, 2020

Saw someone suggest "cooling degree days" (basically the number of days that it's hot enough out that you'd want the AC on) as a proxy for states where it's currently too hot for people to want to be outdoors. It is indeed quite predictive of current COVID spread.

```
regress r_t total_infected cooling_days

      Source |       SS           df       MS      Number of obs   =        51
-------------+----------------------------------   F(2, 48)        =     24.00
       Model |  1.1182775          2  .55913875   Prob > F        =    0.0000
    Residual |  1.11811459        48  .023294054   R-squared       =    0.5000
-------------+----------------------------------   Adj R-squared   =    0.4792
       Total |  2.23639209        50  .044727842   Root MSE        =    .15262

------------------------------------------------------------------------------
         r_t |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
otal_infected |  -.0282222   .0052702    -5.36   0.000    -.0388186   -.0176257
cooling_days |   .0040044   .0010745     3.73   0.001      .001844    .0061649
       _cons |   1.022637   .0425723    24.02   0.000     .9370401    1.108235
------------------------------------------------------------------------------
```

128 870 2,403

Peter Hull
@instrumenthull

Replying to @NateSilver538

Nate. Please man. ", r." Please.

11:05 PM · Jun 21, 2020

Heteroskedasticity-robust standard errors

- Removes assumption that errors all come from the same distribution.

$$\begin{pmatrix} \sigma_1 & & & & \\ 0 & \sigma_2 & & & \\ 0 & 0 & \sigma_3 & & \\ \vdots & \vdots & \vdots & \ddots & \\ 0 & 0 & 0 & \dots & \sigma_N \end{pmatrix}$$

- $se_{het}(\hat{\beta}_1) = \sqrt{\frac{\sum_{i=1}^{N} x_i^2 \hat{u}_i^2}{\sum_{i=1}^{N} x_i^2}}$
- Unless you have a reason to use clustered standard errors, use these.

Heteroskedasticity-robust standard errors

- Estimate regressions with robust standard errors in R using the `feols` function from the package `fixest`.
- `bivariate <- feols(y ~ x, data, vcov = "hetero")`
- Unlike `lm`, you can extract standard errors from a regression estimated with `feols` using `bivariate$se`. Convenient!

Clustered standard errors

- Assume some structure on variance-covariance matrix. For example, errors can be arbitrarily correlated among students in the same school, but are zero otherwise.
- Suppose we cluster within school. Students 1, 2, and 3 attend Boston Latin School; Students 4 and 5 attend Boston Latin Academy:

$$\begin{pmatrix} \sigma_1 & & & & & \\ \sigma_{1,2} & \sigma_2 & & & & \\ \sigma_{1,3} & \sigma_{2,3} & \sigma_3 & & & \\ 0 & 0 & 0 & \sigma_4 & & \\ \vdots & \vdots & \vdots & \vdots & \ddots & \\ 0 & 0 & 0 & \sigma_{4,5} & \dots & \sigma_5 \end{pmatrix}$$

Clustered standard errors

- Clustering often dramatically increases standard errors.
- Cluster at level of treatment. For example, if you want to analyze the relationship between school infrastructure and student achievement, cluster at the school level because all students at a school exposed to the same unexplained school-level factors.
- Cluster on a variable called `school_id` in R by using `fixest` package again.
    - `bivariate <- feols(y ~ x, data, cluster = ~school_id)`.

Which standard errors to use?

You want to know if more generous R&D tax credits are associated with greater R&D spending. Using a sample of firms, you run a regression of R&D spending on the R&D tax credit rate in the firm's state of incorporation.

Cluster at the state level. Firms in the same state have correlated errors, and those errors are likely correlated with the tax credit.

Which standard errors to use?

You want to know if a drug is effective at treating hypertension. You randomly assign a sample of people with hypertension living in the United States to receive the drug or a placebo and run a regression of blood pressure on a dummy variable that indicates whether a person received the drug.

Robust standard errors. People in the same state have correlated errors, but those errors are not likely to be correlated with receiving the drug (which was randomly allocated). General formula for standard error suggests that clustering standard errors should have little impact if X_i and u_i are uncorrelated.

$$se(\hat{\beta}_1) = \sqrt{\frac{Var(\sum_{i=1}^{N} X_i u_i)}{\sum_{i=1}^{N} X_i^2}}$$

Dummy Variables

- Commonly used term for binary variables i.e. 0/1, FALSE/TRUE, No/Yes.
- For example, let $female_i$ be a dummy variable that equals 1 if person i identifies as female and 0 otherwise.

$$wage_i = \beta_0 + \beta_1 female_i + \beta_2 educ_i + u_i.$$

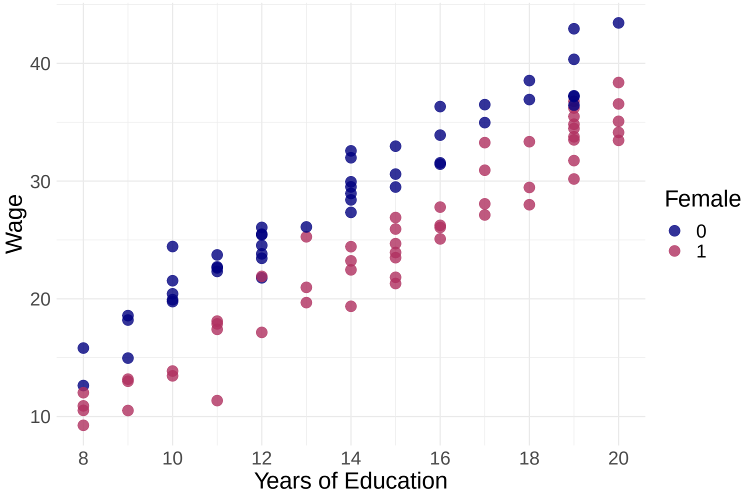
- Suppose $educ_i$ equals i's years of education. Then we can interpret β_1 as the difference in wage between females and non-females, holding years of education constant.
- As we'll discuss more next week, we can create multiple dummy variables to represent categorical variables. For a variable with K categories, use $K - 1$ dummy variables.

Dummy Variables

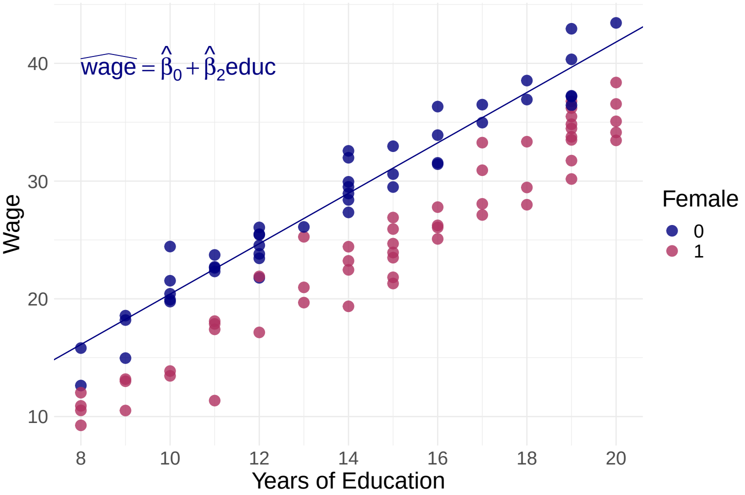
```r
library(tidyverse)
library(fixest)

fit_wage <- feols(wage ~ female + educ, data, vcov = "hetero")
summary(fit_wage)
```

```
OLS estimation, Dep. Var.: wage
Observations: 100
Standard-errors: Heteroskedasticity-robust
            Estimate Std. Error   t value  Pr(>|t|)
(Intercept) -1.01091   0.764066  -1.32306   0.18893
female      -6.16624   0.394913 -15.61418 < 2.2e-16 ***
educ         2.14103   0.053963  39.67614 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 1.9334   Adj. R2: 0.941694
```

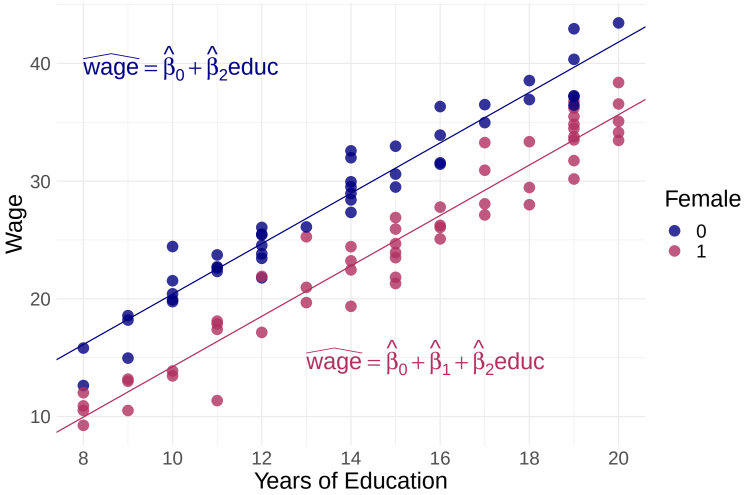
Visualizing Dummy Variables



Visualizing Dummy Variables



Visualizing Dummy Variables



Exercise: NES Data

- In this exercise, I ask the question: do women think more positively than men of having a woman president? If so, can we just attribute this to women having more liberal views?
- Use data from the 2012 National Election Survey, long-running survey of US political attitutes.
- NES asked respondents how they felt about having a woman president in the next 20 years.
- If you follow along, you will need to load the packages `tidyverse`, `haven` and `fixest` as well as dataset `nes2012edit.dta`.

Exercise: NES Data

Let's plan to run the following regressions:

$$\text{swp}_i = \alpha_0 + \alpha_1 \text{female}_i + u_i \tag{4}$$

$$\text{swp}_i = \beta_0 + \beta_1 \text{female}_i + \beta_2 \text{liberal}_i + \beta_3 \text{moderate}_i + u_i \tag{5}$$

Key variables

- `swp`: measure of support for a woman president, see next slide
- `gender`: 1 if male, 2 if female
- `libcon3`: 1 if liberal, 2 if moderate, 3 if conservative

Exercise: NES Data

```r
library(tidyverse)
library(haven)
library(fixest)
# Load data
nes <- read_dta("nes2012edit.dta")
# Print data dictionary
print_labels(nes$swp)
```

```
Labels:
 value                  label
     1         Extremely bad
     2        Moderately bad
     3           A little bad
     4 Neither good nor bad
     5          A little good
     6       Moderately good
     7         Extremely good
```

Exercise: NES Data

1. Drop observations if gender, political ideology, or support for a woman president is missing.
2. Recode gender so that male is 0 and female is 1.
3. Estimate the bivariate regression of `swp` on `female`.
   Use `feols(formula, vcov = "hetero")` to estimate the model with heteroskedasticity-robust standard errors.

```
# Recode data
nes_2 <- nes %>%
  # Drop missing observations
  drop_na(gender, libcon3, swp) %>%
  # Recode gender
  mutate(female = gender - 1)

# Estimate model
bivariate <- feols(swp ~ female, nes_2, vcov = "hetero")
```

Exercise: NES Data

```r
summary(bivariate)
```

```
OLS estimation, Dep. Var.: swp
Observations: 5,324
Standard-errors: Heteroskedasticity-robust
            Estimate Std. Error   t value  Pr(>|t|)
(Intercept) 4.703803   0.027822 169.07019 < 2.2e-16 ***
female      0.385870   0.040485   9.53108 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 1.47892   Adj. R2: 0.016541
```

The standard deviation of `swp` is 1.49, so women's support for a woman president is about 1/4 of a standard deviation higher than men's support on average.

Exercise: NES Data

Estimating a t-test for difference of means yields the SAME t-statistic as the bivariate regression.

```r
t.test(swp ~ female, nes_2)
```

```
    Welch Two Sample t-test

data:  swp by female
t = -9.5311, df = 5316.1, p-value < 2.2e-16
alternative hypothesis: true difference in means between group 0 and group 1 is not
95 percent confidence interval:
 -0.4652376 -0.3065016
sample estimates:
mean in group 0 mean in group 1
       4.703803        5.089673
```

Exercise: NES Data

To evaluate whether the relationship between `female` and `swp` is driven by political ideology, now add the categorical variable `libcon3` which measures political ideology as liberal, moderate, or conservative.

To specify that `libcon3` is a categorical variable, include it in the regression as `i(libcon3, ref = ...)`. Set the base group to "Cons" by replacing `...` with the value of `libcon3` corresponding to "Cons".

```
full_model <- feols(swp ~ female + i(libcon3, ref = 3),
                    data = nes_2,
                    vcov = "hetero")
```

Exercise: NES Data

```r
summary(full_model)
```

```
OLS estimation, Dep. Var.: swp
Observations: 5,324
Standard-errors: Heteroskedasticity-robust
            Estimate Std. Error   t value   Pr(>|t|)
(Intercept) 4.386735   0.033142 132.35992  < 2.2e-16 ***
female      0.338704   0.039322   8.61351  < 2.2e-16 ***
libcon3::1  0.858631   0.044693  19.21162  < 2.2e-16 ***
libcon3::2  0.229783   0.052097   4.41071 1.0507e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 1.42831   Adj. R2: 0.082354
```

Do women think more positively of having a woman president than men after controlling for political ideology? Yes.