

HYPOTHESES TESTING WITH DUMMY VARIABLES

Research question

- Teach for America is a federally-funded program to train non-education majors to be teachers through an intensive program after students graduate with a non-teaching degree. They are often placed in low-income schools that experience teacher shortages.
- We are interested in whether the program is effective with regards to teacher performance in the classroom. Do Teach for America fellows generate better student performance than regular teachers?
- We will compare TFA fellows to other teachers with regular education degrees, and we will control for suburban (typically high-income) and urban (typically low-income) school environments.

Group means as a table:

(group means)	Suburban	Urban
Regular Teachers	75	57
Teach for America	75	66

Average math score in each group
measured in percentiles

Raw data:

Suburban Schools			Urban Schools		
math	d.sub.tfa	d.sub.reg	d.urb.tfa	d.urb.reg	
75	1	0	0	0	TFA
75	1	0	0	0	
75	1	0	0	0	
75	1	0	0	0	
75	0	1	0	0	REG
75	0	1	0	0	
75	0	1	0	0	
75	0	1	0	0	
75	0	1	0	0	
75	0	1	0	0	
66	0	0	1	0	TFA
66	0	0	1	0	
66	0	0	1	0	
66	0	0	1	0	
66	0	0	1	0	
66	0	0	1	0	
57	0	0	0	1	REG
57	0	0	0	1	
57	0	0	0	1	

DV: Math Scores (percentile)

10 Teach for America teaching fellows (tfa)
10 regular teachers (reg)

↕

11 teachers in suburban schools (sub)
9 teachers in urban schools (urb)

10 Teach for America teaching fellows (tfa)
10 regular teachers (reg)



11 teachers in suburban schools (sub)
9 teachers in urban schools (urb)

math	d.sub.tfa	d.sub.reg	d.urb.tfa	d.urb.reg
75	1	0	0	0
75	1	0	0	0
75	1	0	0	0
75	1	0	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
66	0	0	1	0
66	0	0	1	0
66	0	0	1	0
66	0	0	1	0
66	0	0	1	0
66	0	0	1	0
57	0	0	0	1
57	0	0	0	1
57	0	0	0	1

Average performance
SUBURBAN SCHOOLS:

$$\frac{(11)(75)}{11} = 75$$

Average performance
URBAN SCHOOLS:

$$\frac{(6)(66) + (3)(57)}{9} = 63$$

Kids do
better in
suburban
schools

DV: Math Scores (percentile)

10 Teach for America teaching fellows (tfa)
10 regular teachers (reg)

11 teachers in suburban schools (sub)
9 teachers in urban schools (urb)

math	d.sub.tfa	d.sub.reg	d.urb.tfa	d.urb.reg
75	1	0	0	0
75	1	0	0	0
75	1	0	0	0
75	1	0	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
66	0	0	1	0
66	0	0	1	0
66	0	0	1	0
66	0	0	1	0
66	0	0	1	0
66	0	0	1	0
57	0	0	0	1
57	0	0	0	1
57	0	0	0	1

Average performance
TFA INSTRUCTORS:

$$\frac{(4)(75) + (6)(66)}{10} = 69.6$$

Average performance
REGULAR TEACHERS:

$$\frac{(7)(75) + (3)(57)}{10} = 69.6$$

Is Teach for
America
Effective???

No
performance
differences.

Do we trust
these results?
Could they
be biased?

DV: Math Scores (percentile)

10 Teach for America teaching fellows (tfa)
10 regular teachers (reg)



11 teachers in suburban schools (sub)
9 teachers in urban schools (urb)

math	d.sub.tfa	d.sub.reg	d.urb.tfa	d.urb.reg
75	1	0	0	0
75	1	0	0	0
75	1	0	0	0
75	1	0	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
66	0	0	1	0
66	0	0	1	0
66	0	0	1	0
66	0	0	1	0
66	0	0	1	0
66	0	0	1	0
57	0	0	0	1
57	0	0	0	1
57	0	0	0	1



Average performance

TFA INSTRUCTORS IN SUBURBAN SCHOOLS:

$$\frac{(4)(75)}{4} = 75$$

Average performance

REGULAR TEACHERS IN SUBURBAN SCHOOLS:

$$\frac{(7)(75)}{7} = 75$$

Performance of both teacher types
is identical in suburban schools



DV: Math Scores (percentile)

10 Teach for America teaching fellows (tfa)
10 regular teachers (reg)



11 teachers in suburban schools (sub)
9 teachers in urban schools (urb)

math	d.sub.tfa	d.sub.reg	d.urb.tfa	d.urb.reg
75	1	0	0	0
75	1	0	0	0
75	1	0	0	0
75	1	0	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
66	0	0	1	0
66	0	0	1	0
66	0	0	1	0
66	0	0	1	0
66	0	0	1	0
66	0	0	1	0
57	0	0	0	1
57	0	0	0	1
57	0	0	0	1



Teach for America is effective training
for teachers in urban schools.

Average performance
TFA INSTRUCTORS IN SUBURBAN SCHOOLS:

$$\frac{(6)(66)}{6} = 66$$

Average performance
REGULAR TEACHERS IN URBAN SCHOOLS:

$$\frac{(3)(57)}{3} = 57$$

9-point
performance
difference in
urban schools
for teacher
types!

↑
DV: Math Scores (percentile)

Question:

10 Teach for America teaching fellows (tfa)
10 regular teachers (reg)

11 teachers in suburban schools (sub)
9 teachers in urban schools (urb)



	Suburban Schools		Urban Schools	
math	d.sub.tfa	d.sub.reg	d.urb.tfa	d.urb.reg
75	1	0	0	0
75	1	0	0	0
75	1	0	0	0
75	1	0	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
66	0	0	1	0
66	0	0	1	0
66	0	0	1	0
66	0	0	1	0
66	0	0	1	0
66	0	0	1	0
57	0	0	0	1
57	0	0	0	1
57	0	0	0	1

TFA
REG
TFA
REG

Average performance
TFA INSTRUCTORS:

$$\frac{(4)(75) + (6)(66)}{10} = 69.6$$

Average performance
REGULAR TEACHERS:

$$\frac{(7)(75) + (3)(57)}{10} = 69.6$$

Why do we have no performance difference when comparing teaching programs directly?

DV: Math Scores (percentile)

10 Teach for America teaching fellows (tfa)
10 regular teachers (reg)



11 teachers in suburban schools (sub)
9 teachers in urban schools (urb)

math	d.sub.tfa	d.sub.reg	d.urb.tfa	d.urb.reg
75	1	0	0	0
75	1	0	0	0
75	1	0	0	0
75	1	0	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
66	0	0	1	0
66	0	0	1	0
66	0	0	1	0
66	0	0	1	0
66	0	0	1	0
66	0	0	1	0
57	0	0	0	1
57	0	0	0	1
57	0	0	0	1



But we find a 9-point
difference here?

Average performance
TFA INSTRUCTORS IN SUBURBAN SCHOOLS:

$$\frac{(6)(66)}{6} = 66 \quad \leftarrow$$

Average performance
REGULAR TEACHERS IN URBAN SCHOOLS:

$$\frac{(3)(57)}{3} = 57 \quad \leftarrow$$

↑
DV: Math Scores (percentile)

10 Teach for America teaching fellows (tfa)
10 regular teachers (reg)



11 teachers in suburban schools (sub)
9 teachers in urban schools (urb)

math	d.sub.tfa	d.sub.reg	d.urb.tfa	d.urb.reg
75	1	0	0	0
75	1	0	0	0
75	1	0	0	0
75	1	0	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
66	0	0	1	0
66	0	0	1	0
66	0	0	1	0
66	0	0	1	0
66	0	0	1	0
66	0	0	1	0
57	0	0	0	1
57	0	0	0	1
57	0	0	0	1

Notice the selection process:

Teach for America instructors more likely to teach in URBAN schools
(6 out of 10)

Regular instructors are more likely to select SUBURBAN schools
(7 out of 10)

DV: Math Scores (percentile)

10 Teach for America teaching fellows (tfa)
10 regular teachers (reg)



11 teachers in suburban schools (sub)
9 teachers in urban schools (urb)

math	d.sub.tfa	d.sub.reg	d.urb.tfa	d.urb.reg
75	1	0	0	0
75	1	0	0	0
75	1	0	0	0
75	1	0	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
66	0	0	1	0
66	0	0	1	0
66	0	0	1	0
66	0	0	1	0
66	0	0	1	0
66	0	0	1	0
57	0	0	0	1
57	0	0	0	1
57	0	0	0	1



DV: Math Scores (percentile)

Average performance
TFA INSTRUCTORS:

$$\frac{(4)(75) + (6)(66)}{10} = 69.6$$

Average performance
REGULAR TEACHERS:

$$\frac{(7)(75) + (3)(57)}{10} = 69.6$$

So comparing
teaching programs
without controlling for
differences in teaching
environments leads to
bias.

We incorrectly
conclude the TFA
training program is **NOT**
working when **IT IS**, but
just in urban schools.

Specification: Dummy Variable Design Matrix

10 Teach for America teaching fellows (tfa)
10 regular teachers (reg)



11 teachers in suburban schools (sub)
9 teachers in urban schools (urb)

math	d.sub.tfa	d.sub.reg	d.urb.tfa	d.urb.reg
75	1	0	0	0
75	1	0	0	0
75	1	0	0	0
75	1	0	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
75	0	1	0	0
66	0	0	1	0
66	0	0	1	0
66	0	0	1	0
66	0	0	1	0
66	0	0	1	0
66	0	0	1	0
57	0	0	0	1
57	0	0	0	1
57	0	0	0	1



Average performance
TFA IN SUBURBAN SCHOOLS:



REGULAR TEACHERS IN SUBURBAN SCHOOLS:



TFA IN URBAN SCHOOLS:



REGULAR TEACHERS IN URBAN SCHOOLS:



DV: Math Scores (percentile)

10 Teach for America teaching fellows (tfa)
10 regular teachers (reg)



11 teachers in suburban schools (sub)
9 teachers in urban schools (urb)

math	d.sub.tfa	d.sub.reg	d.urb.tfa	d.urb.reg	D1+D2+D3+D4	Intercept
75	1	0	0	0	1	1
75	1	0	0	0	1	1
75	1	0	0	0	1	1
75	1	0	0	0	1	1
75	0	1	0	0	1	1
75	0	1	0	0	1	1
75	0	1	0	0	1	1
75	0	1	0	0	1	1
75	0	1	0	0	1	1
75	0	1	0	0	1	1
75	0	1	0	0	1	1
66	0	0	1	0	1	1
66	0	0	1	0	1	1
66	0	0	1	0	1	1
66	0	0	1	0	1	1
66	0	0	1	0	1	1
66	0	0	1	0	1	1
57	0	0	0	1	1	1
57	0	0	0	1	1	1
57	0	0	0	1	1	1

This is a fully-interacted design matrix where there is exactly one dummy variable for each group.

Since the linear combination of all dummy variables would give us a columns of 1's, **we cannot run this model with an intercept due to perfect multi-collinearity** (you cannot include two identical variables in a model – the statistics program will automatically drop one.

`lm( math ~ d.sub.tfa + d.sub.reg + d.urb.tfa + d.urb.reg )`

(group means)	Suburban	Urban
Regular Teachers	75	57
Teach for America	75	66

(design matrix)

math	d.sub.tfa	d.sub.reg	d.urb.tfa	d.urb.reg
75	1	0	0	0
75	0	1	0	0
66	0	0	1	0
57	0	0	0	1

$$\text{math} = \overset{(75)}{b_1} \cdot \text{d.sub.tfa} + \overset{(75)}{b_2} \cdot \text{d.sub.reg} + \overset{(66)}{b_3} \cdot \text{d.urb.tfa} + \overset{(57)}{b_4} \cdot \text{d.urb.reg}$$

Each coefficient represents a separate group mean. Note, there is no intercept!

R code for dropping the intercept: add a **-1** at the end.

`lm( math ~ d.sub.tfa + d.sub.reg + d.urb.tfa + d.urb.reg - 1 )`

(group means)	Suburban	Urban
Regular Teachers	75	57
Teach for America	75	66

(design matrix construction)

```
d.sub <- ifelse( school == "suburban", 1, 0 )
d.reg <- ifelse( teacher == "regular", 1, 0 )
d.sub.reg <- d.sub * d.reg
```

```
lm( math ~ d.sub + d.reg + d.sub.reg )
```

Urban schools and TFA teachers are both omitted.
So d.urb.tfa becomes the reference group b0.

(design matrix)

math	intercept	d.sub	d.reg	d.sub.reg
75	1	1	0	0
75	1	1	1	1
66	1	0	0	0
57	1	0	1	0

$$\text{math} = b_0 + b_1 \cdot d.\text{sub} + b_2 \cdot d.\text{reg} + b_3 \cdot d.\text{sub.reg}$$

The groups are now additive.

$$b_0 + b_1 = 66 + 9 = 75 \quad (\text{suburban TFA})$$

$$b_0 + b_1 + b_2 + b_3 = 66 + 9 - 9 + 9 = 75 \quad (\text{suburban regular})$$

$$b_0 = 66 \quad (\text{urban TFA} - \text{reference group})$$

$$b_0 + b_2 = 66 - 9 = 57 \quad (\text{urban regular})$$

(group means)	Suburban	Urban
Regular Teachers	75	57
Teach for America	75	66

(design matrix)

math	d.sub.tfa	d.sub.reg	d.urb.tfa	d.urb.reg
75	1	0	0	0
75	0	1	0	0
66	0	0	1	0
57	0	0	0	1

(design matrix)

math	intercept	d.sub	d.reg	d.sub.reg
75	1	1	0	0
75	1	1	1	1
66	1	0	0	0
57	1	0	1	0

$$\text{math} = \overset{(75)}{b_1} \cdot \text{d.sub.tfa} + \overset{(75)}{b_2} \cdot \text{d.sub.reg} + \overset{(66)}{b_3} \cdot \text{d.urb.tfa} + \overset{(57)}{b_4} \cdot \text{d.urb.reg}$$

Each coefficient represents a separate group mean. Note, there is no intercept!

$$\text{math} = \overset{(66)}{b_0} + \overset{(9)}{b_1} \cdot \text{d.sub} + \overset{(-9)}{b_2} \cdot \text{d.reg} + \overset{(9)}{b_3} \cdot \text{d.sub.reg}$$

The groups are now additive.

$$b_0 + b_1 = 66 + 9 = \mathbf{75} \quad (\text{suburban TFA})$$

$$b_0 + b_1 + b_2 + b_3 = 66 + 9 - 9 + 9 = \mathbf{75} \quad (\text{suburban regular})$$

$$b_0 = \mathbf{66} \quad (\text{urban TFA} - \text{reference group})$$

$$b_0 + b_2 = 66 - 9 = \mathbf{57} \quad (\text{urban regular})$$

`lm( math ~ d.sub + d.reg + d.sub.reg )`

(group means)	Suburban	Urban
Regular Teachers	75	57
Teach for America	75	66

(design matrix)

math	intercept	d.urb	d.tfa	d.urb.tfa
75	1	0	1	0
75	1	0	0	0
66	1	1	1	1
57	1	1	0	0

(design matrix)

math	intercept	d.sub	d.reg	d.sub.reg
75	1	1	0	0
75	1	1	1	1
66	1	0	0	0
57	1	0	1	0

No matter which groups you omit, you can always recover the group means. You just multiply all coefficients by the appropriate row in the design matrix.

$$\begin{matrix} (75) & (-18) & (0) & (9) \\ \text{math} = b_0 + b_1 \cdot \text{d.urb} + b_2 \cdot \text{d.tfa} + b_3 \cdot \text{d.urb.tfa} \end{matrix}$$

$$b_0(1) + b_1(0) + b_2(1) + b_3(0) = 75 + 0 = \mathbf{75} \quad (\text{suburban TFA})$$

$$b_0(1) + b_1(0) + b_2(0) + b_3(0) = \mathbf{75} \quad (\text{suburban regular} - \text{ref. group})$$

$$b_0(1) + b_1(1) + b_2(1) + b_3(1) = 75 - 18 + 0 + 9 = \mathbf{66} \quad (\text{urban TFA})$$

$$b_0(1) + b_1(1) + b_2(0) + b_3(0) = 75 - 18 = \mathbf{57} \quad (\text{urban regular})$$

$$\begin{matrix} (66) & (9) & (-9) & (9) \\ \text{math} = b_0 + b_1 \cdot \text{d.sub} + b_2 \cdot \text{d.reg} + b_3 \cdot \text{d.sub.reg} \end{matrix}$$

$$b_0(1) + b_1(1) + b_2(0) + b_3(0) = 66 + 9 = \mathbf{75} \quad (\text{suburban TFA})$$

$$b_0(1) + b_1(1) + b_2(1) + b_3(1) = 66 + 9 - 9 + 9 = \mathbf{75} \quad (\text{suburban regular})$$

$$b_0(1) + b_1(0) + b_2(0) + b_3(0) = \mathbf{66} \quad (\text{urban TFA} - \text{ref. group})$$

$$b_0(1) + b_1(0) + b_2(1) + b_3(0) = 66 - 9 = \mathbf{57} \quad (\text{urban regular})$$

Hypothesis-Testing

(group means)	Suburban	Urban
Regular Teachers	75	57
Teach for America	75	66

(design matrix)

math	d.sub.tfa	d.sub.reg	d.urb.tfa	d.urb.reg
75	1	0	0	0
75	0	1	0	0
66	0	0	1	0
57	0	0	0	1

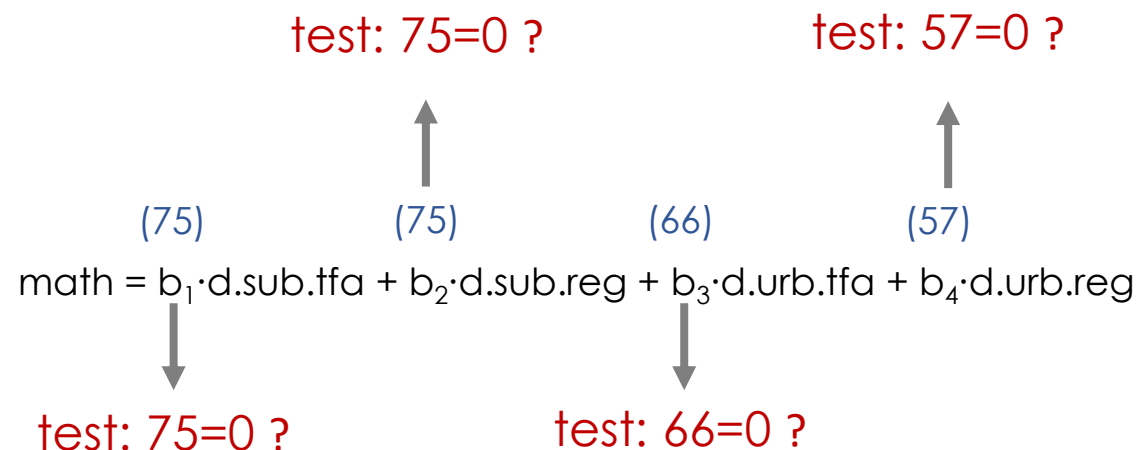
$$\text{math} = \overset{(75)}{b_1} \cdot \text{d.sub.tfa} + \overset{(75)}{b_2} \cdot \text{d.sub.reg} + \overset{(66)}{b_3} \cdot \text{d.urb.tfa} + \overset{(57)}{b_4} \cdot \text{d.urb.reg}$$

If this is the most intuitive way to get group means, why don't we run this regression model?

(group means)	Suburban	Urban
Regular Teachers	75	57
Teach for America	75	66

(design matrix)

math	d.sub.tfa	d.sub.reg	d.urb.tfa	d.urb.reg
75	1	0	0	0
75	0	1	0	0
66	0	0	1	0
57	0	0	0	1

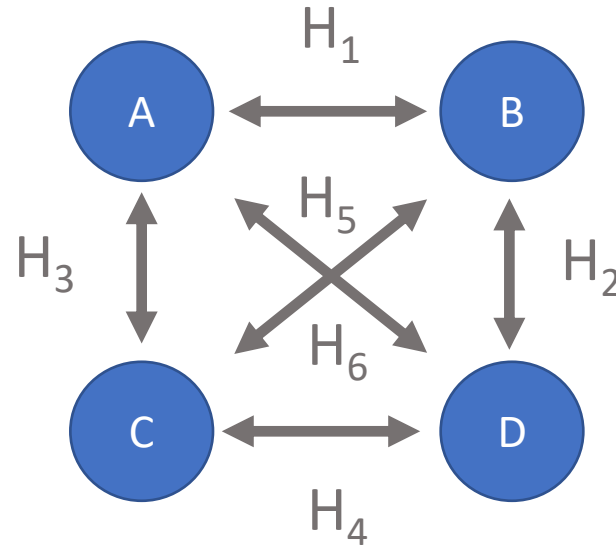


Our research question is whether teachers trained in the Teach for America (TFA) program perform better than teachers trained in the regular program? None of the tests for significance of coefficients b_0 to b_3 in this regression reflect meaningful tests. We already know the group means will not be zero!

It makes it easier to calculate group means, but makes it impossible to answer our research question based upon regression results.

ALL POSSIBLE TESTS (CONTRASTS)

(group means)	Suburban	Urban
Regular Teachers	A	B
Teach for America	C	D



H_1 : $A = B$? Do regular teachers perform differently in urban and suburban schools?

H_2 : $B = D$? Do regular and TFA teachers perform different in urban schools?

H_3 : $A = C$? Do regular and TFA teachers perform different in suburban schools?

etc...

(group means)	Suburban	Urban
Regular Teachers	75	57
Teach for America	75	66

(design matrix)

math	intercept	d.sub	d.reg	d.sub.reg
75	1	1	0	0
75	1	1	1	1
66	1	0	0	0
57	1	0	1	0

$$\text{math} = b_0 + b_1 \cdot \text{d.sub} + b_2 \cdot \text{d.reg} + b_3 \cdot \text{d.sub.reg}$$

The groups are now additive.

$$b_0 + b_1 = 66 + 9 = 75 \quad (\text{suburban TFA})$$

$$b_0 + b_1 + b_2 + b_3 = 66 + 9 - 9 + 9 = 75 \quad (\text{suburban regular})$$

$$b_0 = 66 \quad (\text{urban TFA} - \text{reference group})$$

$$b_0 + b_2 = 66 - 9 = 57 \quad (\text{urban regular})$$

$$\begin{aligned} b_0 &= b_0 + b_1 \\ 0 &= b_1 \end{aligned} \quad (\text{suburban TFA different than urban TFA?})$$

$$\begin{aligned} b_0 &= b_0 + b_2 \\ 0 &= b_2 \end{aligned} \quad (\text{urban regular different than urban TFA?})$$

$$\begin{aligned} b_0 + b_1 + b_2 &= b_0 + b_1 + b_2 + b_3 \\ 0 &= b_3 \end{aligned} \quad (\text{suburban } \times \text{ regular different than suburban + regular ?})$$

Each specification creates a set of hypotheses tests.

We can never test all hypotheses with a single model, but we get several tests from one.

(group means)	Suburban	Urban
Regular Teachers	A	B
Teach for America	C	D

(design matrix)

math	intercept	d.sub	d.reg	d.sub.reg
75	1	1	0	0
75	1	1	1	1
66	1	0	0	0
57	1	0	1	0

(66) (9) (-9) (9)

$$\text{math} = b_0 + b_1 \cdot \text{d.sub} + b_2 \cdot \text{d.reg} + b_3 \cdot \text{d.sub.reg}$$

$$b_0 + b_1 = 66 + 9 = 75 \quad (\text{suburban TFA})$$

$$b_0 + b_1 + b_2 + b_3 = 66 + 9 - 9 + 9 = 75 \quad (\text{suburban regular})$$

$$b_0 = 66 \quad (\text{urban TFA} - \text{reference group})$$

$$b_0 + b_2 = 66 - 9 = 57 \quad (\text{urban regular})$$

$$b_0 = b_0 + b_1$$

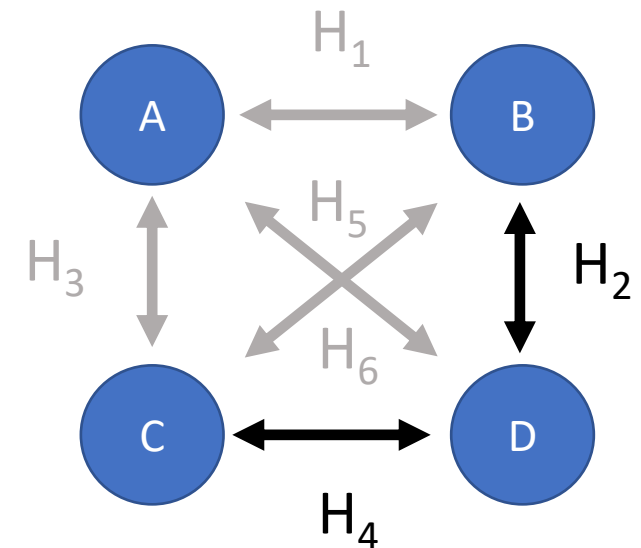
$$0 = b_1 \quad (\text{suburban TFA different than urban TFA?})$$

$$b_0 = b_0 + b_2$$

$$0 = b_2 \quad (\text{urban regular different than urban TFA?})$$

$$b_0 + b_1 + b_2 = b_0 + b_1 + b_2 + b_3$$

$$0 = b_3 \quad (\text{suburban} \times \text{regular different than suburban + regular ?})$$



(group means)	Suburban	Urban
Regular Teachers	A	B
Teach for America	C	D

(design matrix)

math	intercept	d.urb	d.tfa	d.urb.tfa
75	1	0	1	0
75	1	0	0	0
66	1	1	1	1
57	1	1	0	0

$$\text{math} = b_0 + b_1 \cdot \text{d.urb} + b_2 \cdot \text{d.tfa} + b_3 \cdot \text{d.urb.tfa}$$

(75) (-18) (0) (9)

$$b_0 + b_2 = 75 + 0 = 75 \quad (\text{suburban TFA})$$

$$b_0 = 75 \quad (\text{suburban regular} - \text{reference group})$$

$$b_0 + b_1 = 75 - 18 + 9 = 66 \quad (\text{urban TFA})$$

$$b_0 + b_2 = 75 - 18 = 57 \quad (\text{urban regular})$$

$$b_0 = b_0 + b_1$$

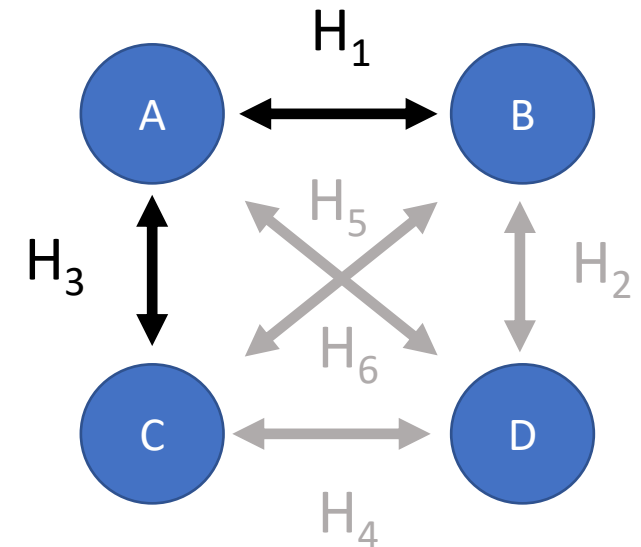
$$0 = b_1 \quad (\text{suburban regular different than urban regular?})$$

$$b_0 = b_0 + b_2$$

$$0 = b_2 \quad (\text{suburban regular different than suburban TFA?})$$

$$b_0 + b_1 + b_2 = b_0 + b_1 + b_2 + b_3$$

$$0 = b_3 \quad (\text{urban } \times \text{ TFA different than urban + TFA ?})$$

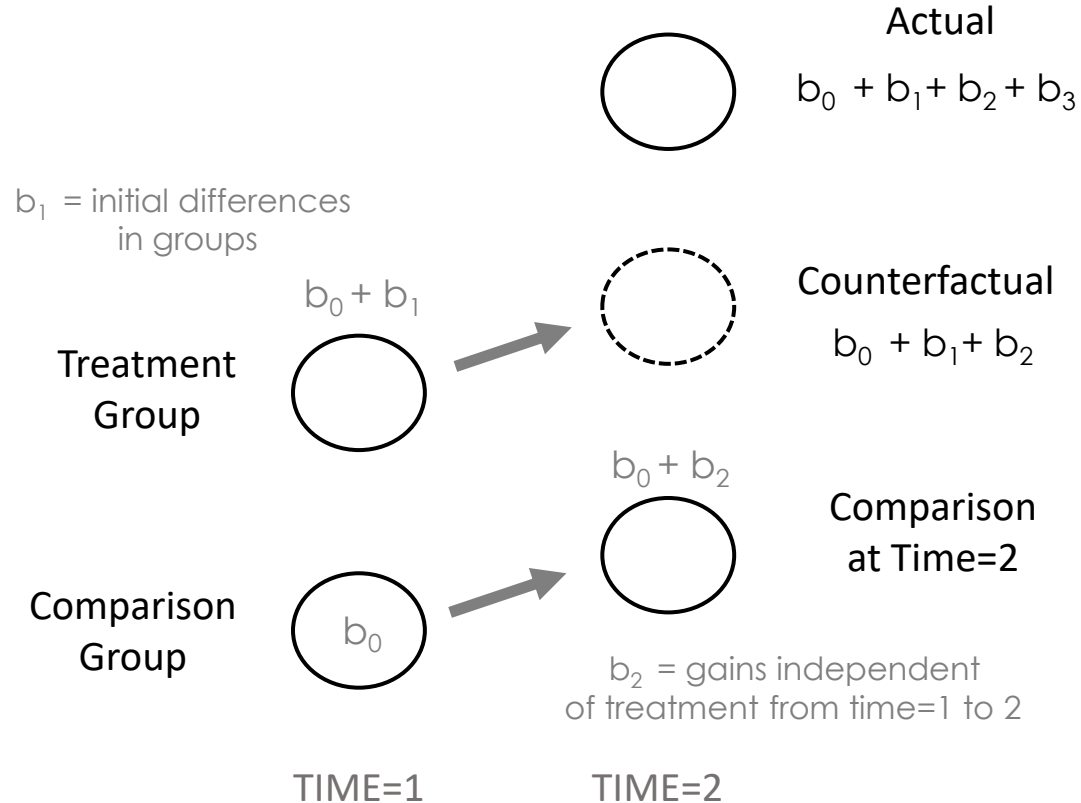


Test for treatment effects in pre-post design with control group:

Does the treatment group improve more than expected ?
(the counterfactual captures the expectation if they have similar gains as control)

$$\text{outcome} = b_0 + b_1 \cdot \text{d.treat} + b_2 \cdot \text{d.time2} + b_3 \cdot \text{d.treat.post}$$

$$b_0 + b_1 + b_2 = b_0 + b_1 + b_2 + b_3$$
$$0 = b_3$$



Expected treatment group mean if program is ineffective

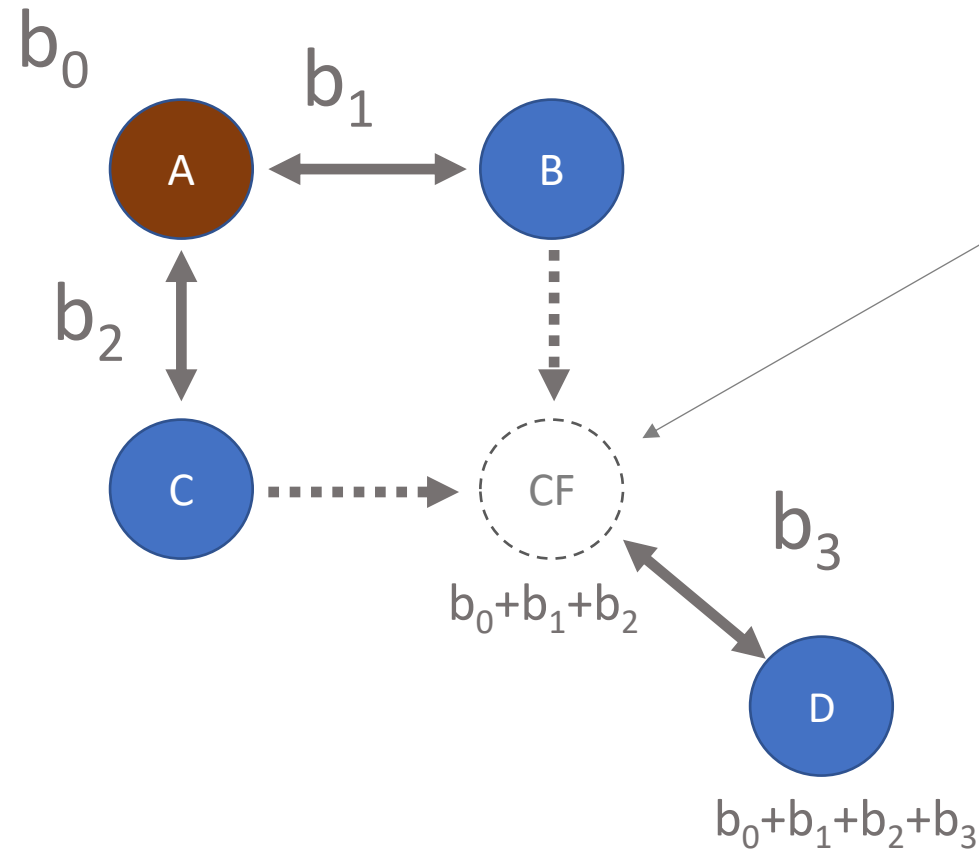
Actual observed group mean

$$b_0 + b_1 + b_2 = b_0 + b_1 + b_2 + b_3$$

$$0 = b_3 ?$$

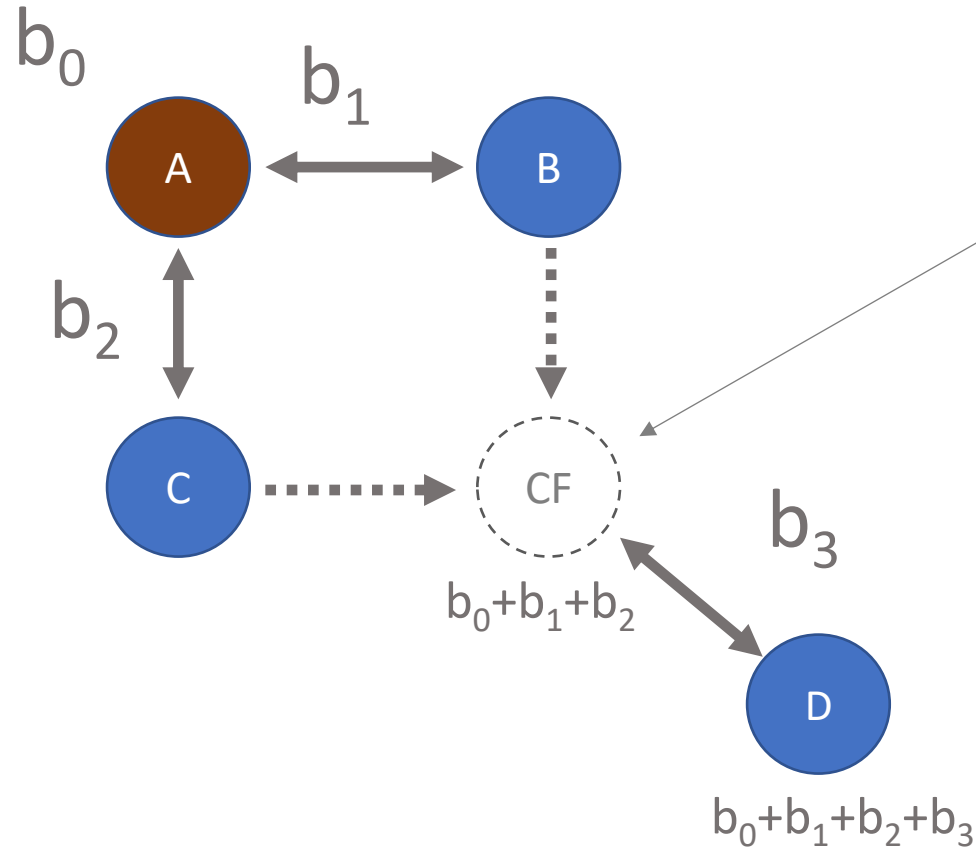
b_3 : Test for whether the treatment was effective – if the group looks different than we would expect

Test for b_3 : $b_0 + b_1 + b_2 = b_0 + b_1 + b_2 + b_3$



The counterfactual is where we expect D to be if there is no independent effect of being in the group BxC.

Test for b_3 : $b_0 + b_1 + b_2 = b_0 + b_1 + b_2 + b_3$



BxC might be something like race+sex. If minorities are paid less than whites in the population, and if women are paid less than men in the population, then will the penalty for a female minority employee be additive (minority penalty + female penalty = CF) or is there an EXTRA penalty for being female AND minority? The additional penalty is picked up by the minority_x_female interaction. The coefficient b_3 is a test of whether the penalty is simply additive ($D=CF :: b_3=0$) or if minority women are paid even less than expected if race and gender were independent ($b_3 \neq 0$).

Example, **does a blood pressure medication work the same for men and women?** Men and women in the study are randomly assigned to treatment and control groups. Measures are all at the end of the study after 6 months.

$$BP = b_0 + b_1 \cdot d.treat + b_2 \cdot d.female + b_3 \cdot d.treat.female$$

Group means:

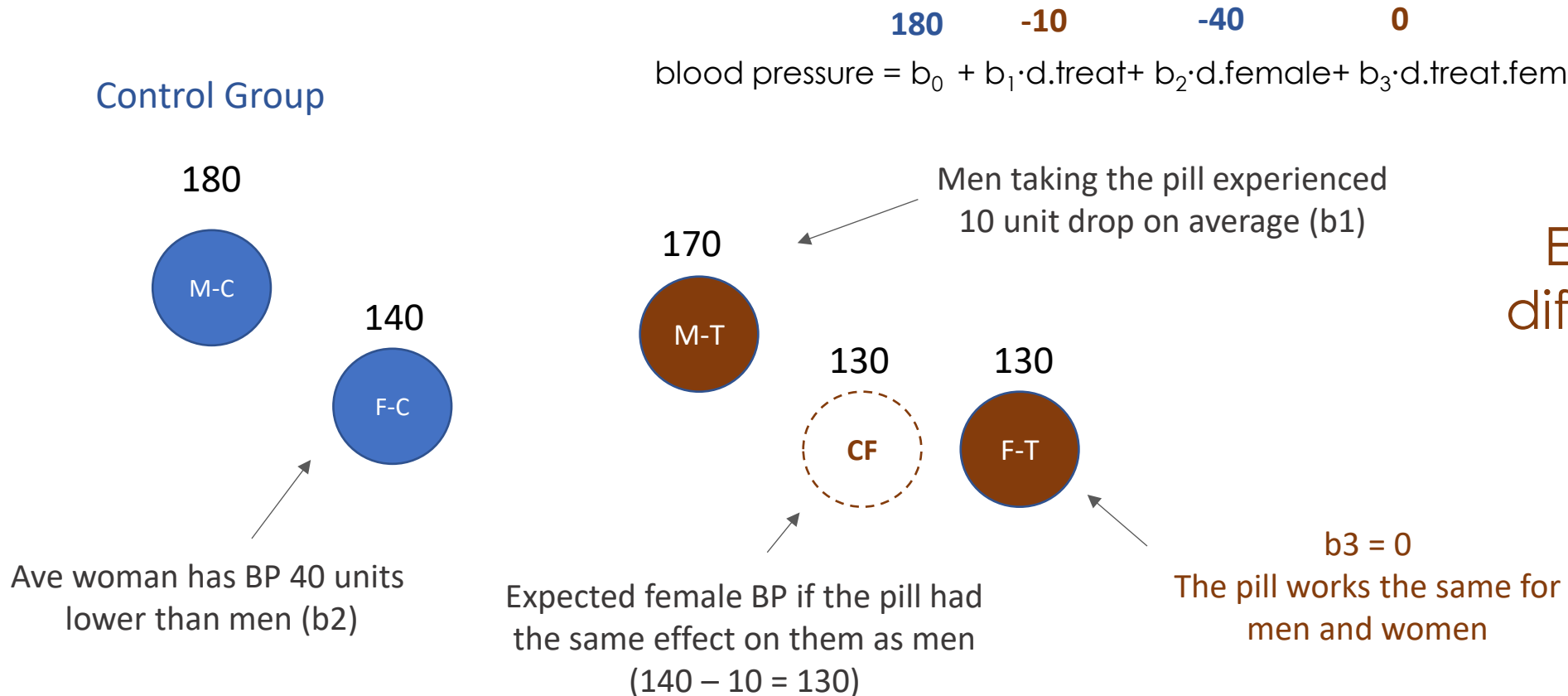
b_0 = final blood pressure of men in control group (receive placebo pill)

$b_0 + b_1$ = final blood pressure of men in the treatment group

$b_0 + b_2$ = final blood pressure of women in the control group

$b_0 + b_1 + b_2$ = final blood pressure of women in the treatment group if gains M/F equal

$b_0 + b_1 + b_2 + b_3$ = final blood pressure of women in treatment if gains M/F not equal



Example of **NO** differential effects for men and women

$b_3 = 0$
The pill works the same for men and women

Example, **does a BP pill work the same for men and women?** Men and women in the study are randomly assigned to treatment and control groups. Measures are all at the end of the study after 6 months.

$$BP = b_0 + b_1 \cdot d.treat + b_2 \cdot d.female + b_3 \cdot d.treat.female$$

Group means:

b_0 = final BP of men in control group (receive placebo pill)

$b_0 + b_1$ = final BP of men in the treatment group

$b_0 + b_2$ = final BP of women in the control group

$b_0 + b_1 + b_2$ = final BP of women in the treatment group if gains M/F equal

$b_0 + b_1 + b_2 + b_3$ = final BP of women in treatment if gains M/F not equal

180

-10

-40

0

$$weight = b_0 + b_1 \cdot d.treat + b_2 \cdot d.female + b_3 \cdot d.treat.female$$



180

-10

-40

$$BP = b_0 + b_1 \cdot d.treat + b_2 \cdot d.female$$

Example of **NO differential effects** for men and women

Note, if we determine that there are no differences in treatment effect size for men and women then we can use the simpler model. Final BP of women in the treatment group is just $b_0 + b_1 + b_2$ since $b_3 = 0$

Example, **does a BP pill work the same for men and women?** Men and women in the study are randomly assigned to treatment and control groups. Measures are all at the end of the study after 6 months.

$$BP = b_0 + b_1 \cdot d.treat + b_2 \cdot d.female + b_3 \cdot d.treat.female$$

Group means:

b_0 = final BP of men in control group (receive placebo pill)

$b_0 + b_1$ = final BP of men in the treatment group

$b_0 + b_2$ = final BP of women in the control group

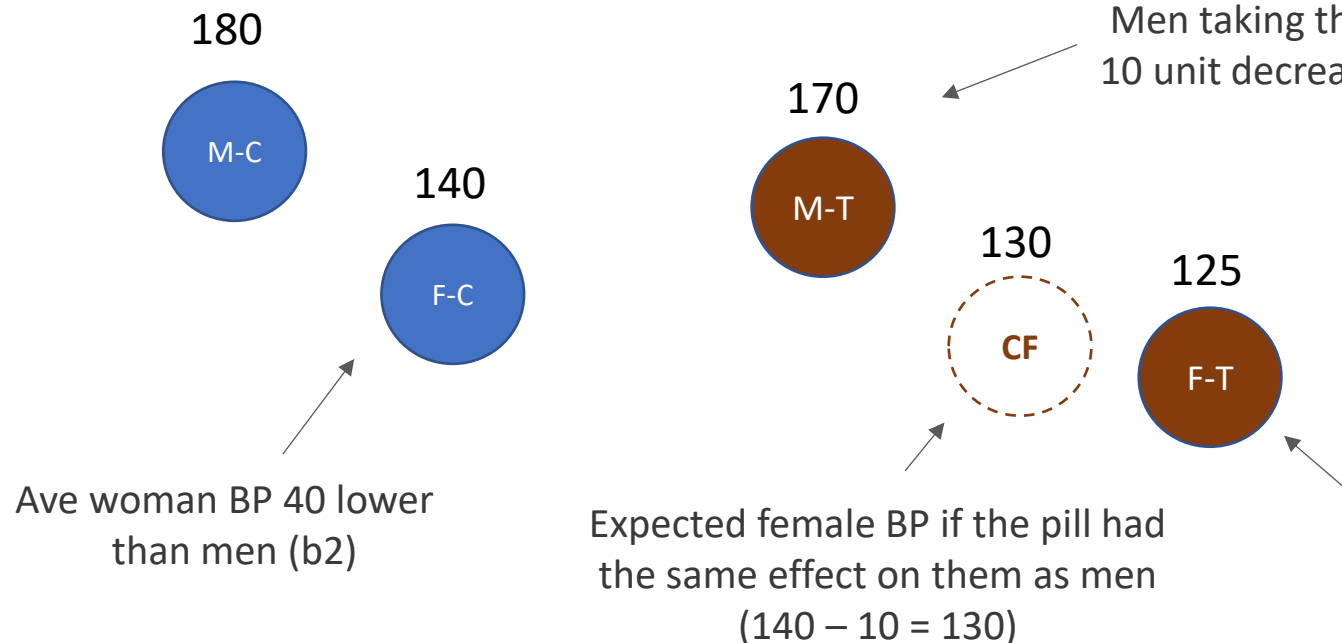
$b_0 + b_1 + b_2$ = final BP of women in the treatment group if gains M/F equal

$b_0 + b_1 + b_2 + b_3$ = final BP of women in treatment if gains M/F not equal

$$180 \quad -10 \quad -40 \quad -5$$

$$BP = b_0 + b_1 \cdot d.treat + b_2 \cdot d.female + b_3 \cdot d.treat.female$$

Control Group



Example of differential effects for men and women

$$b_3 = -5$$

The pill has a larger effect on women than men

Example, **does a diet pill work the same for men and women?** Men and women in the study are randomly assigned to treatment and control groups. Measures are all at the end of the study after 6 months.

$$\text{weight} = b_0 + b_1 \cdot d.\text{treat} + b_2 \cdot d.\text{female} + b_3 \cdot d.\text{treat} \cdot d.\text{female}$$

Group means:

b_0 = final weight of men in control group (receive placebo pill)

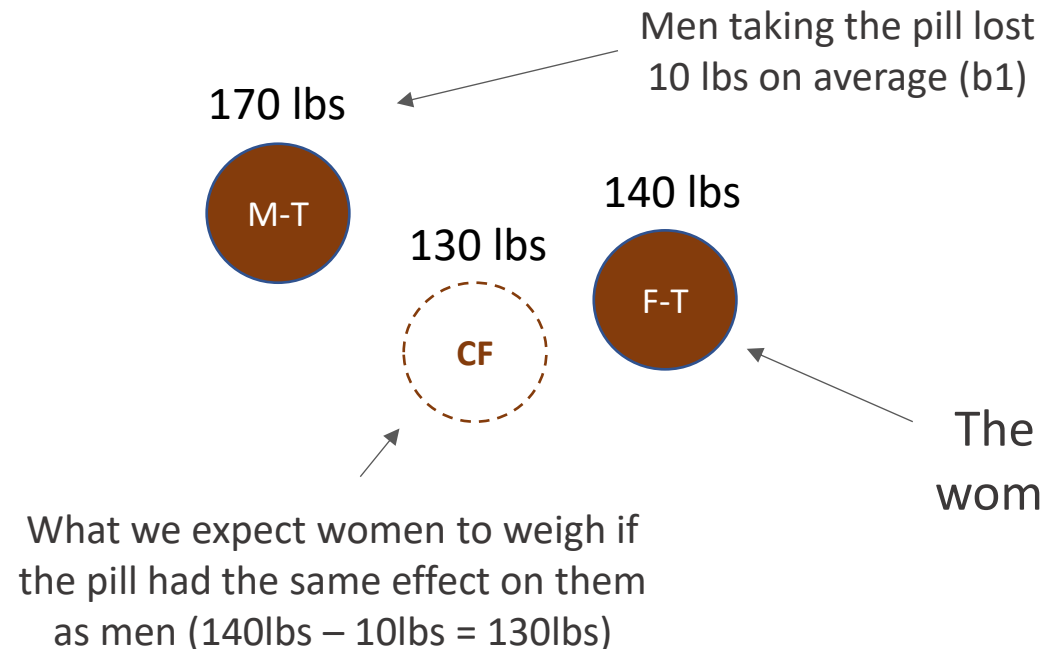
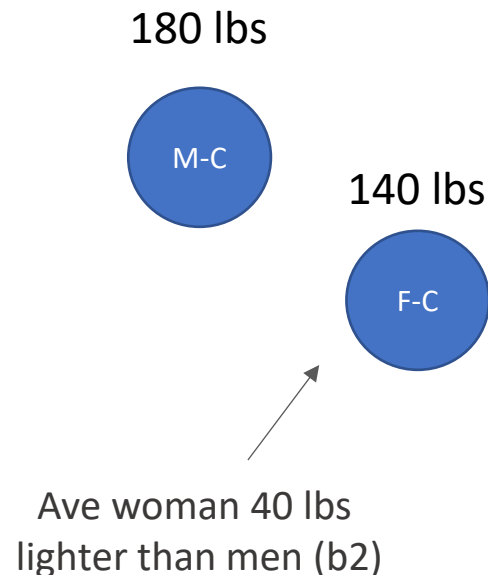
$b_0 + b_1$ = final weight of men in the treatment group

$b_0 + b_2$ = final weight of women in the control group

$b_0 + b_1 + b_2$ = final weight of women in the treatment group if gains M/F equal

$b_0 + b_1 + b_2 + b_3$ = final weight of women in treatment if gains M/F not equal

Control Group



180 - 10 - 40 10

$$\text{weight} = b_0 + b_1 \cdot d.\text{treat} + b_2 \cdot d.\text{female} + b_3 \cdot d.\text{treat} \cdot d.\text{female}$$

Example of differential effects for men and women

$b_3 = 10$

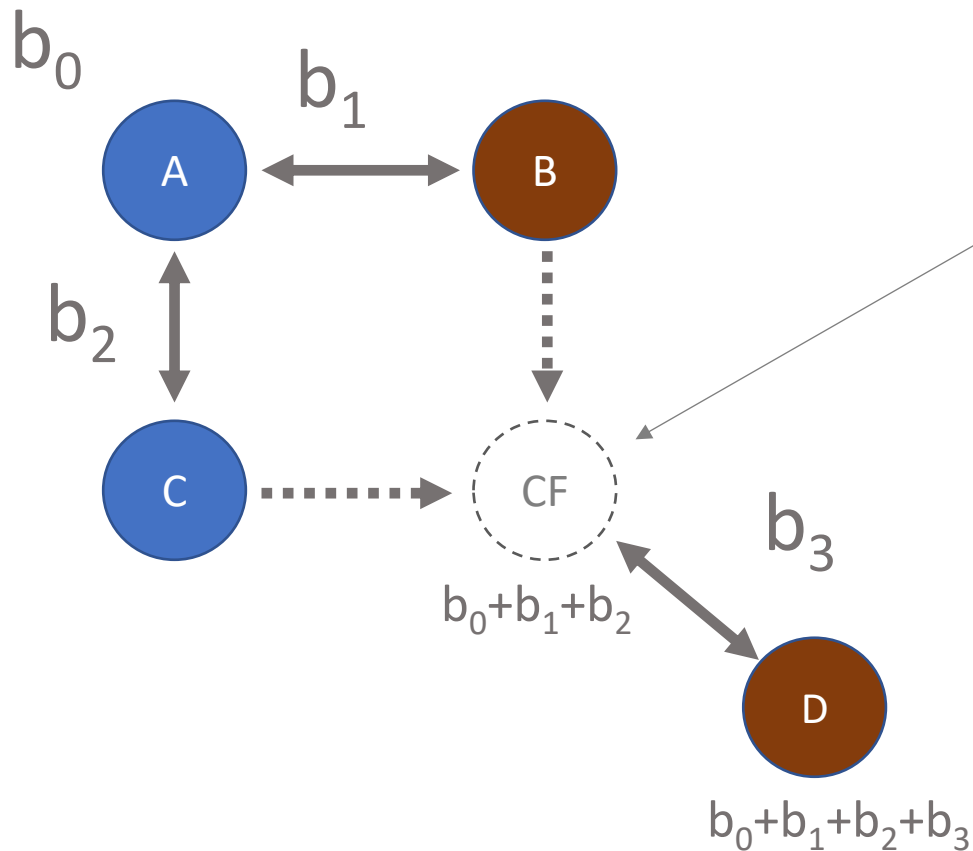
The pill has no effect on women (b_1 and b_3 cancel each other out)

Model:

$$\text{BP} = b_0 + b_1 \cdot \text{d.treat} + b_2 \cdot \text{d.female} + b_3 \cdot \text{d.treat} \cdot \text{d.female}$$

180 - 10 - 40 - 5

Test for b_3 : $b_0 + b_1 + b_2 = b_0 + b_1 + b_2 + b_3$



To create the counterfactual we start we use the female dummy to adjust for differences between men and women, and we use the treatment effect experienced by men to determine the final BP of women if they respond to the medication the same as men.

The interaction $\text{treat} \cdot \text{female}$ tells us if women respond differently to the treatment ($b_3 \neq 0$).

(group means)	Suburban	Urban
Regular Teachers	A	B
Teach for America	C	D

(design matrix)

math	intercept	d.urb	d.tfa	d.urb.tfa
75	1	0	1	0
75	1	0	0	0
66	1	1	1	1
57	1	1	0	0

Back to the TFT example:

$$\text{math} = b_0 + b_1 \cdot d.\text{urb} + b_2 \cdot d.\text{tfa} + b_3 \cdot d.\text{urb.tfa}$$

(75) (-18) (0) (9)

C: $b_0 + b_2 = 75 + 0 = 75$ (suburban TFA)

A: $b_0 = 75$ (suburban regular – **reference group**)

D: $b_0 + b_1 + b_2 + b_3 = 75 - 18 + 0 + 9 = 66$ (urban TFA)

B: $b_0 + b_2 = 75 - 18 = 57$ (urban regular)

If we use regular suburban teachers as the reference group that allows us to answer the questions in our study because the two hypotheses we are about are:

Do TFA teachers perform better in suburban settings?

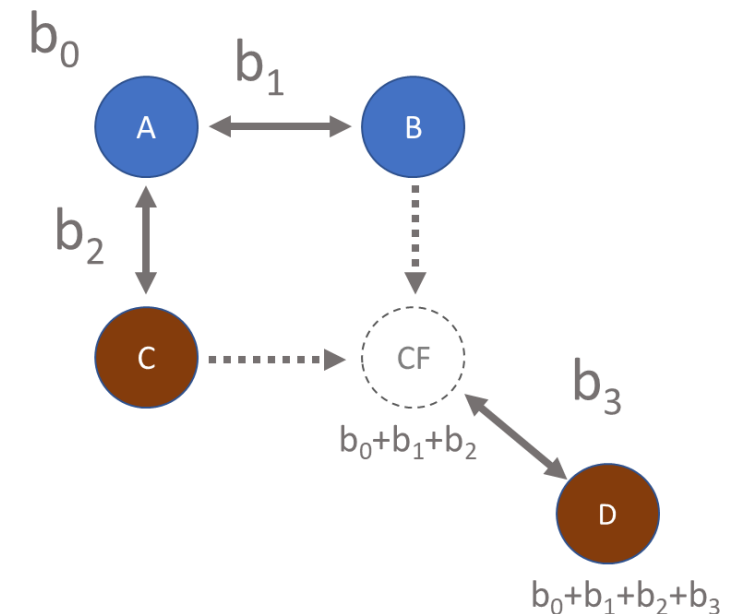
Do TFA teachers perform better in urban settings?

Note that by using the urban and TFA dummy variables we get:

b_2 = tests whether TFA teachers do better in suburban schools

b_3 = tests whether TFA teachers do better than expected in urban schools (expected meaning they perform at the same deficit as regular teachers)

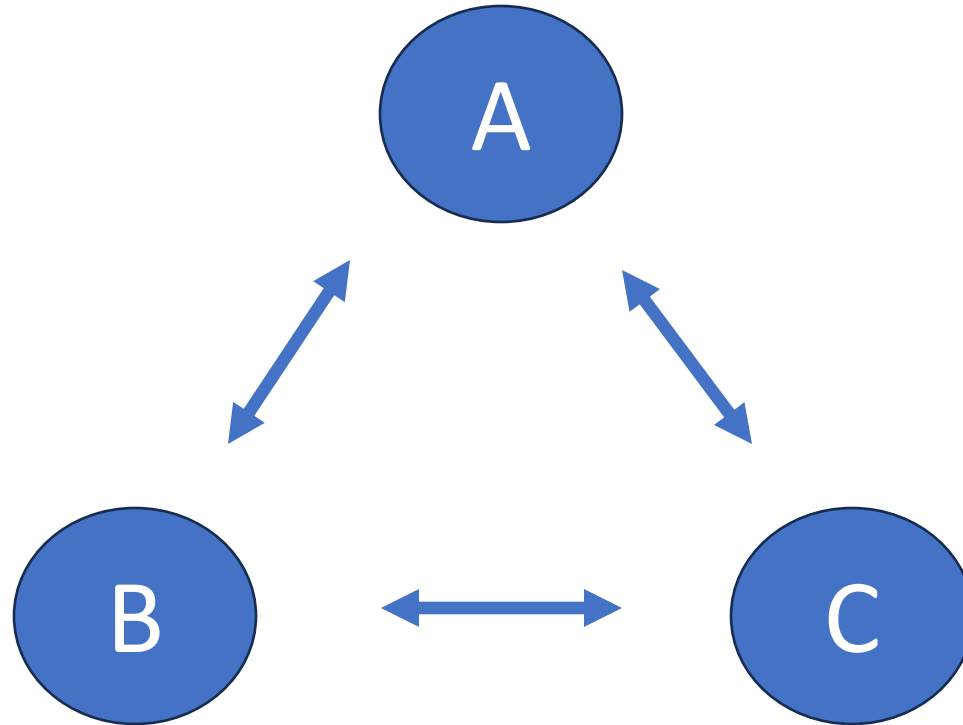
b_3 is not a direct test of group means $B=D$. It's a test of $D=\text{counterfactual } D$ (performance of TFA teachers if no differences in urban schools)



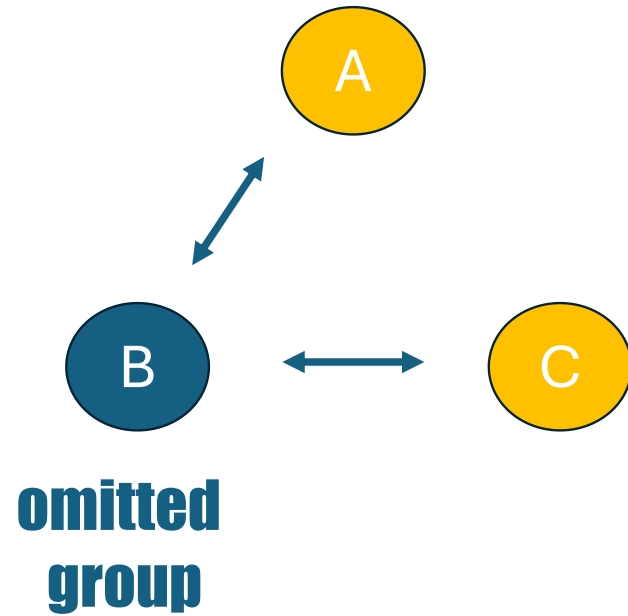
Number of Groups

Factor with 3 levels becomes:

- **Dummy_A**
- **Dummy_B**
- **Dummy_C**



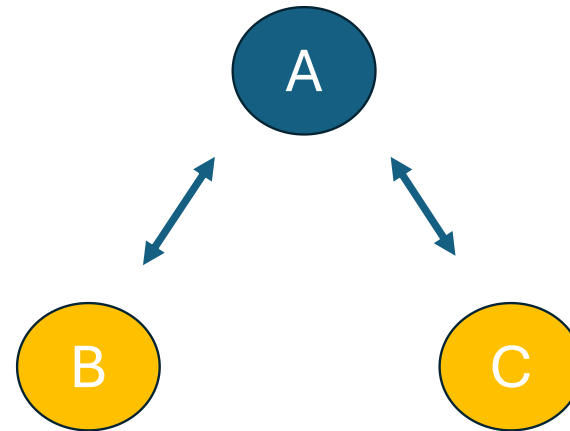
Factor with 3 levels



$$Y = b_0 + b_1 \text{ Dummy_A} + b_2 \text{ Dummy_C}$$

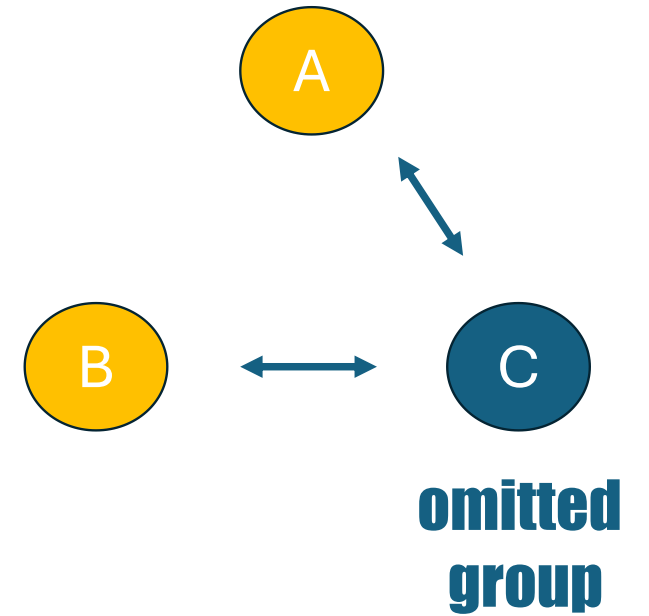
B

omitted group

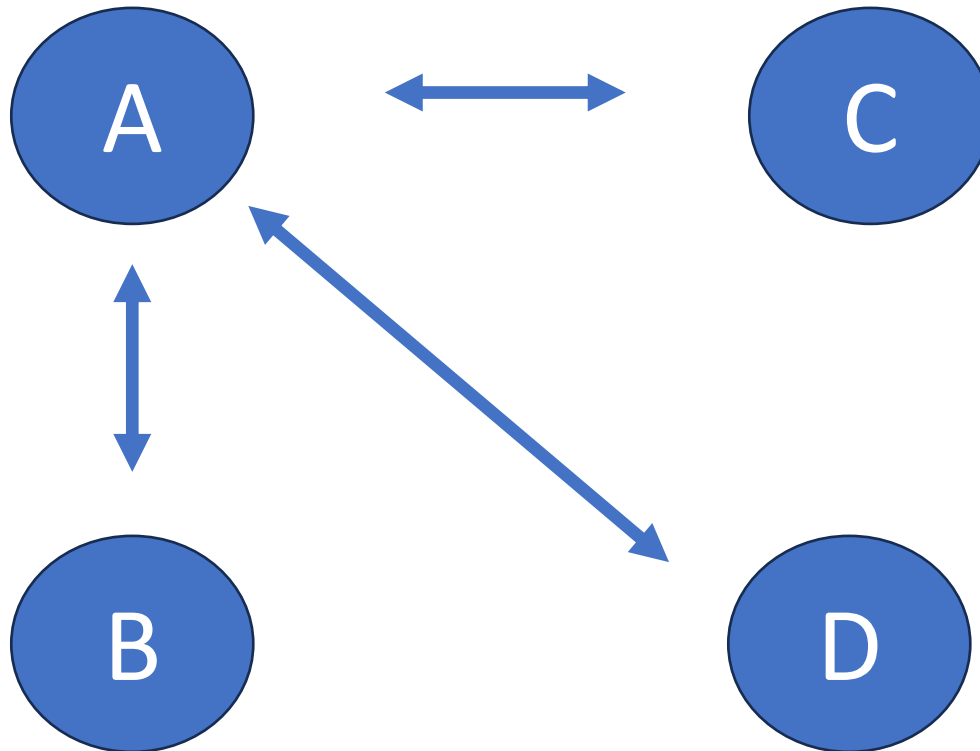


C

$$Y = b_0 + b_1 \text{ Dummy_A} + b_2 \text{ Dummy_B}$$



Factor with 4 levels



Why don't interactions make sense after you turn a categorical variable into dummies?

Why don't
interactions make
sense after you
turn a categorical
variable into
dummies?

White_dummy	Black_dummy
1	0
1	0
0	1
0	1

Group Variable in Venn Diagram

