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	
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
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
SUBURBAN SCHOOLS:

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

Average performance
URBAN SCHOOLS:

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

Kids do better in suburban schools

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**

DV: Math Scores (percentile)

Average performance
TFA INSTRUCTORS:

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

Average performance
REGULAR TEACHERS:

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

Is Teach for America Effective???

No performance differences.

Do we trust these results? Could they be biased?

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**

DV: Math Scores (percentile)

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

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**

DV: Math Scores (percentile)

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!

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	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	
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)

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?

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**

DV: Math Scores (percentile)

But we find a 9-point difference here?

Average performance
TFA INSTRUCTORS IN SUBURBAN SCHOOLS:

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

Average performance
REGULAR TEACHERS IN URBAN SCHOOLS:

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

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**

DV: Math Scores (percentile)

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)

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**

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.

(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**

(75) (75) (66) (57)

$$math = b_1 \cdot d.sub.tfa + b_2 \cdot d.sub.reg + b_3 \cdot d.urb.tfa + b_4 \cdot d.urb.reg$$

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

(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)**

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

The groups are now additive.

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

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

$b_0 = \mathbf{66}$ (urban TFA – **reference group**)

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

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

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.

(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

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

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

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

$b_0 = \mathbf{75}$ (suburban regular – **reference group**)

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

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

(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)

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

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

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

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

$b_0 + b_2 = 66 - 9 = \mathbf{57}$ (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**

(75) (75) (66) (57)

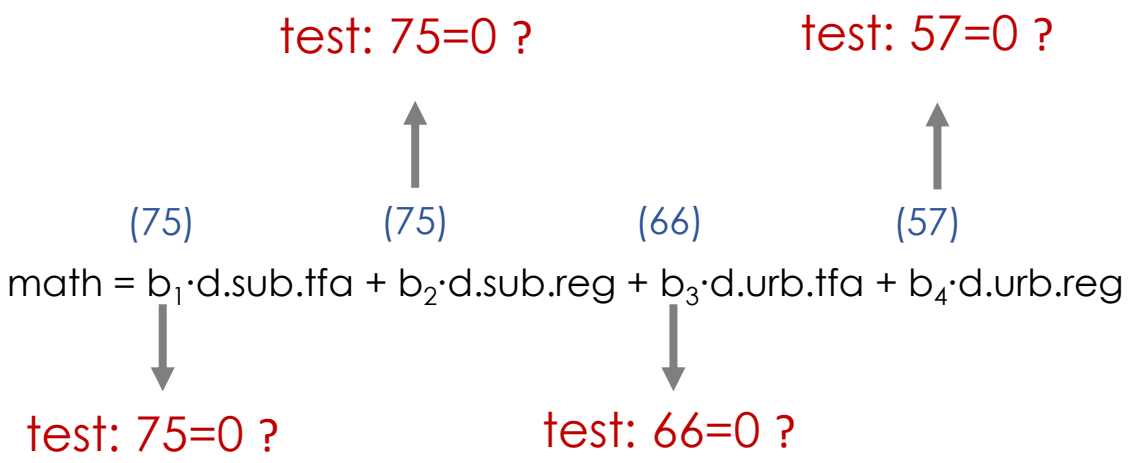
$$math = b_1 \cdot d.sub.tfa + b_2 \cdot d.sub.reg + b_3 \cdot d.urb.tfa + b_4 \cdot d.urb.reg$$

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**

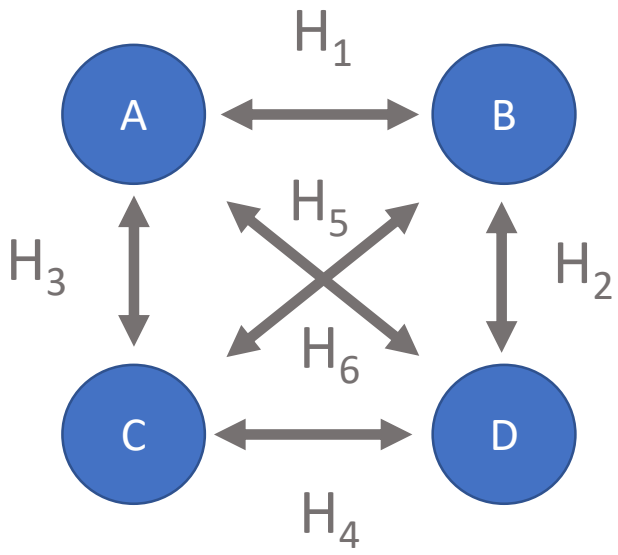


Our research question is whether teachers trained in the Teach for America (TFA) program perform better than teachers trained in the regular program? None of the tests for significance of coefficients b_0 to b_3 in this regression reflect meaningful tests. We already know the group means will not be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

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

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

The groups are now additive.

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

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

$b_0 = \mathbf{66}$ (urban TFA – **reference group**)

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

$b_0 = b_0 + b_1$
$0 = b_1$ (suburban TFA different than urban TFA?)

$b_0 = b_0 + b_2$
$0 = b_2$ (urban regular different than urban TFA?)

$b_0 + b_1 + b_2 = b_0 + b_1 + b_2 + b_3$
$0 = b_3$
(suburban **x** 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)

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

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

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

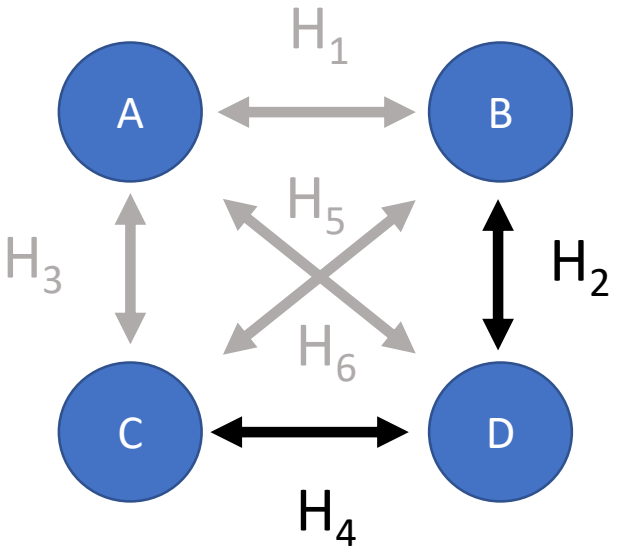
$b_0 = \mathbf{66}$ (urban TFA – **reference group**)

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

$b_0 = b_0 + b_1$
$0 = b_1$ (suburban TFA different than urban TFA?)

$b_0 = b_0 + b_2$
$0 = b_2$ (urban regular different than urban TFA?)

$b_0 + b_1 + b_2 = b_0 + b_1 + b_2 + b_3$
$0 = b_3$
(suburban **x** 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

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

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

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

$b_0 = \mathbf{75}$ (suburban regular – **reference group**)

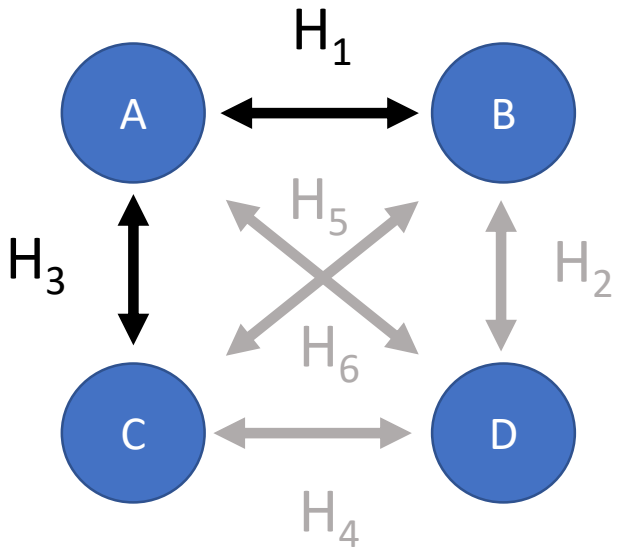
$b_0 + b_1 = 75 - 18 + 9 = \mathbf{66}$ (urban TFA)

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

$b_0 = b_0 + b_1$
$0 = b_1$ (suburban regular different than urban regular?)

$b_0 = b_0 + b_2$
$0 = b_2$ (suburban regular different than suburban TFA?)

$b_0 + b_1 + b_2 = b_0 + b_1 + b_2 + b_3$
$0 = b_3$
(urban **x** TFA different than urban + TFA ?)

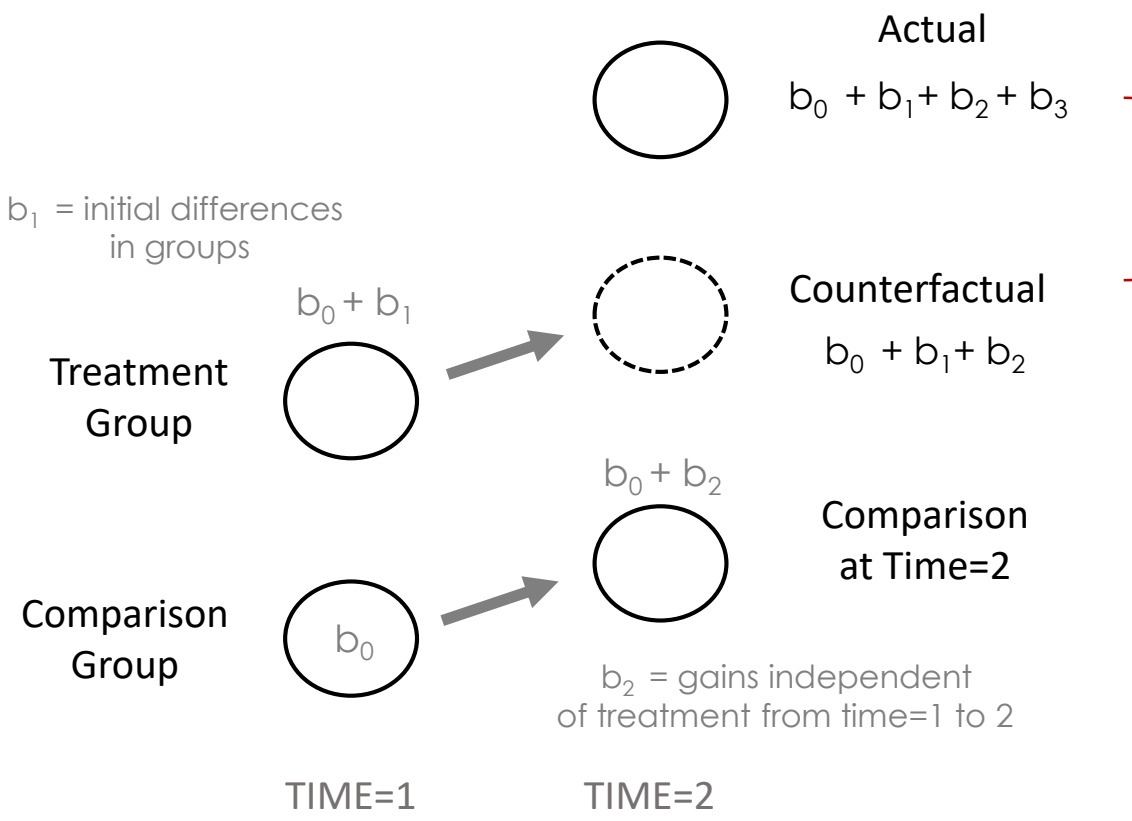


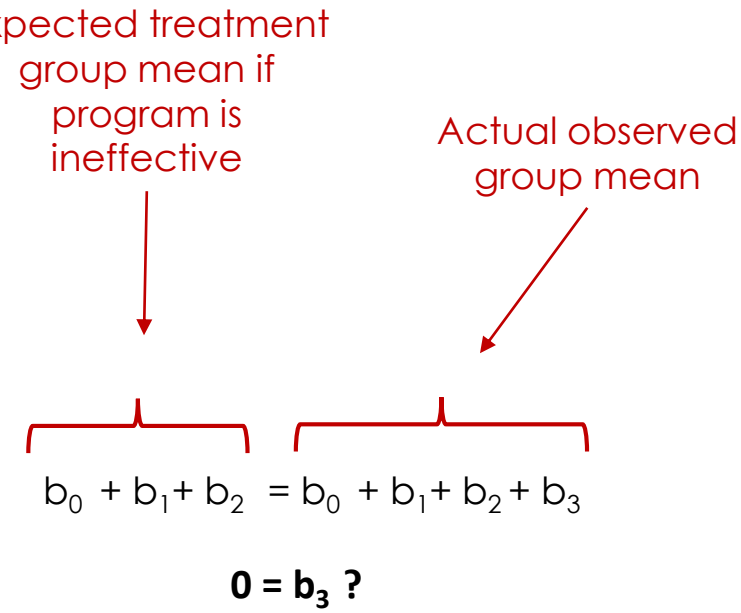
Test for treatment effects in pre-post design with control group:

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

outcome = b_0 + b_1·d.treat+ b_2·d.time2+ b_3·d.treat.post

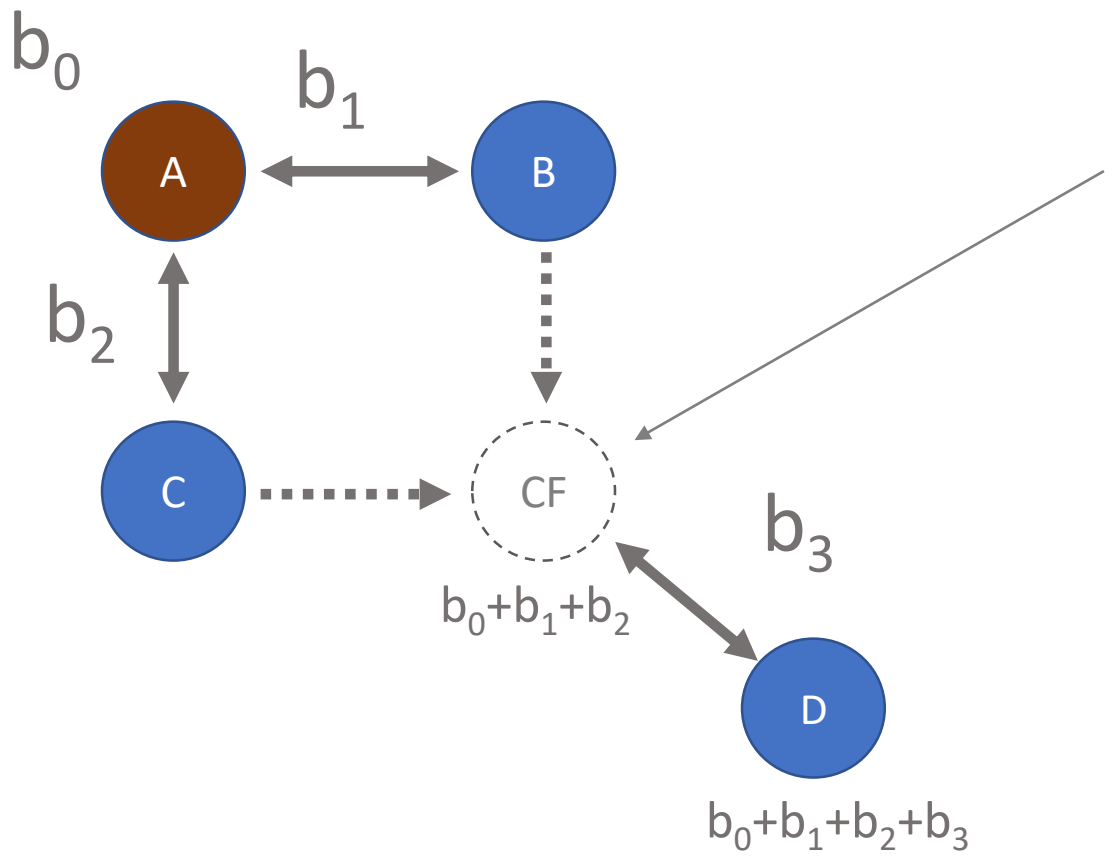
b_0 + b_1+ b_2 = b_0 + b_1+ b_2 + b_3
0 = b_3





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 D1*D2.

Example, **does a diet pill work <u>the same</u> 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.

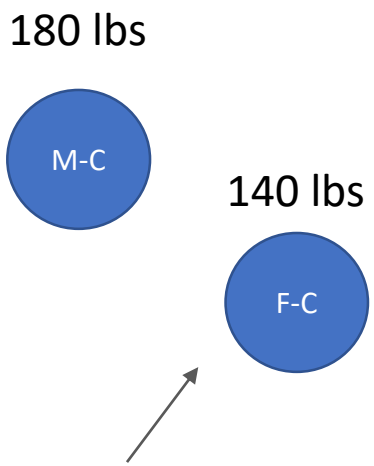
$weight = b_0 + b_1 \cdot d.treat + b_2 \cdot d.female + b_3 \cdot d.treat.female$

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

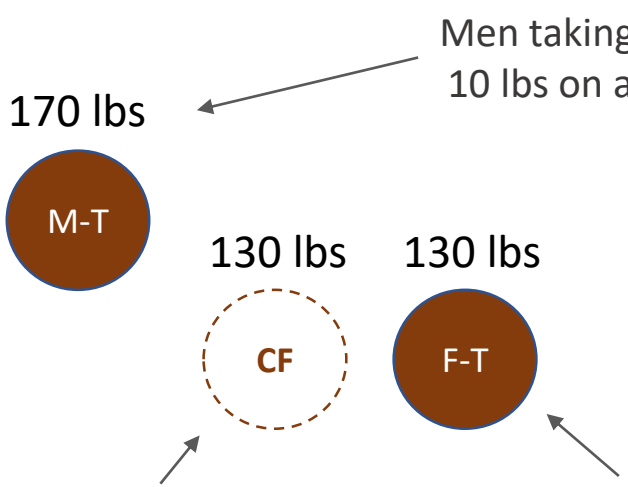
180 **-10** **-40** **0**

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

Control Group



Ave woman 40 lbs lighter than men (b2)



Men taking the pill lost 10 lbs on average (b1)

What we expect women to weigh if the pill had the same effect on them as men (140lbs – 10lbs = 130lbs)

b3 = 0
The pill works the same for men and women

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

Example, **does a diet pill work <u>the same</u> 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.

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

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

180 -10 -40 0

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

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

180 -10 -40

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

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 weight of women in the treatment group is just $b_0+b_1+b_2$ since $b_3=0$

Example, **does a diet pill work <u>the same</u> 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.

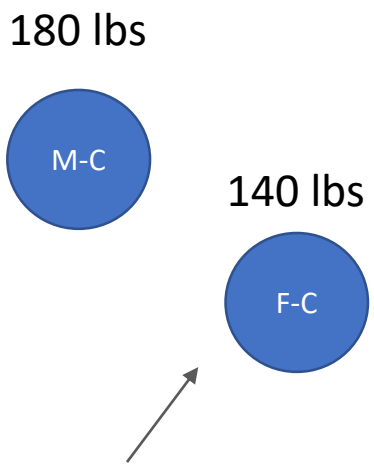
$weight = b_0 + b_1 \cdot d.treat + b_2 \cdot d.female + b_3 \cdot d.treat.female$

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

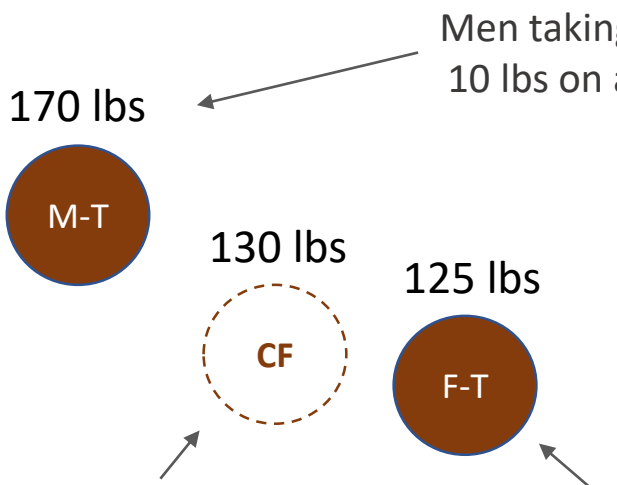
180 – 10 – 40 – 5

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

Control Group



Ave woman 40 lbs lighter than men (b2)



Men taking the pill lost 10 lbs on average (b1)

What we expect women to weigh if the pill had the same effect on them as men (140lbs – 10lbs = 130lbs)

Example of differential effects for men and women

b3 = – 5

The pill has a large effect on women than on men

Example, **does a diet pill work <u>the same</u> 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.

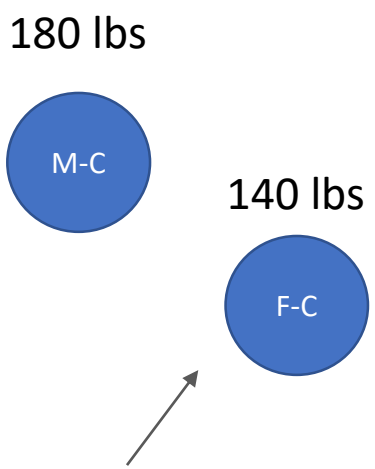
$weight = b_0 + b_1 \cdot d.treat + b_2 \cdot d.female + b_3 \cdot d.treat.female$

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

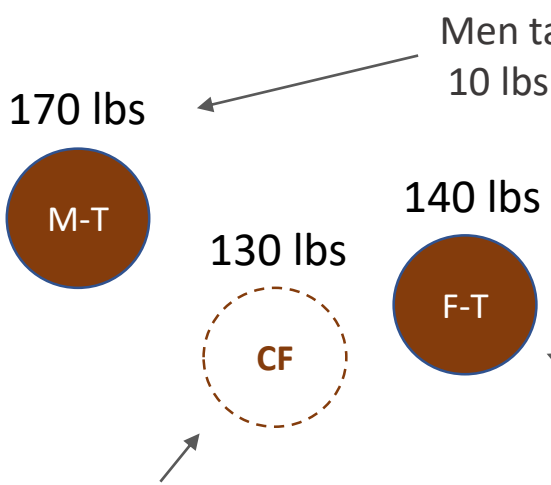
180 **– 10** **– 40** **10**

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

Control Group



Ave woman 40 lbs lighter than men (b2)



What we expect women to weigh if the pill had the same effect on them as men (140lbs – 10lbs = 130lbs)

Example of differential effects for men and women

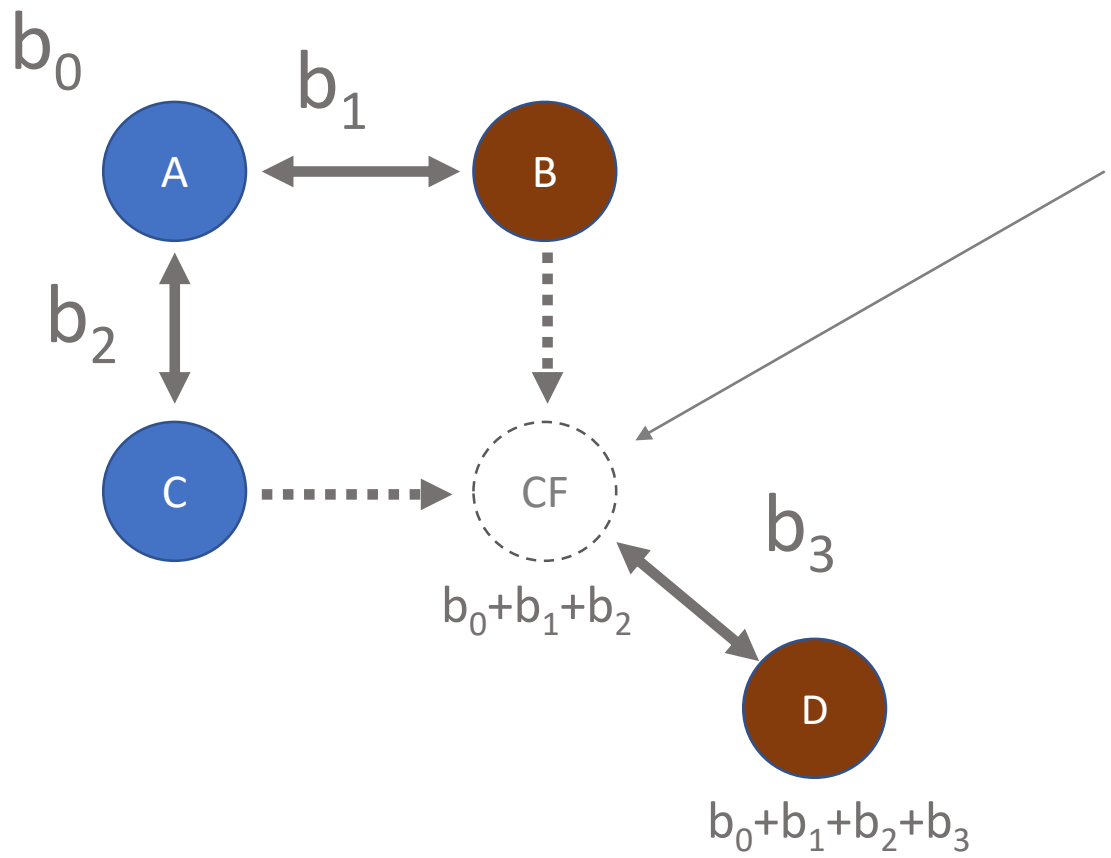
b3 = 10

The pill has no effect on women (b1 and b3 cancel each other out)

Model: $\underset{180}{} \quad \underset{-10}{} \quad \underset{-40}{} \quad \underset{-5}{}$

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

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 weight differences between men and women, and we use the treatment effect of men to determine the final weight of women if they respond the same to the diet pill.

The interaction treat*female tells us if women respond differently to the treatment.

(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:

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

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

C: $b_0 + b_2 = 75 + 0 = \mathbf{75}$ (suburban TFA)

A: $b_0 = \mathbf{75}$ (suburban regular – **reference group**)

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

B: $b_0 + b_2 = 75 - 18 = \mathbf{57}$ (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:

b2 = tests whether TFA teachers do better in suburban schools
b3 = tests whether TFA teachers do better than expected in urban schools (expected meaning they perform at the same deficit as regular teachers)

b3 is not a direct test of group means B=D. It's a test of D=counterfactual D (performance of TFA teachers if no differences in urban schools)