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

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)

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

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:

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

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)

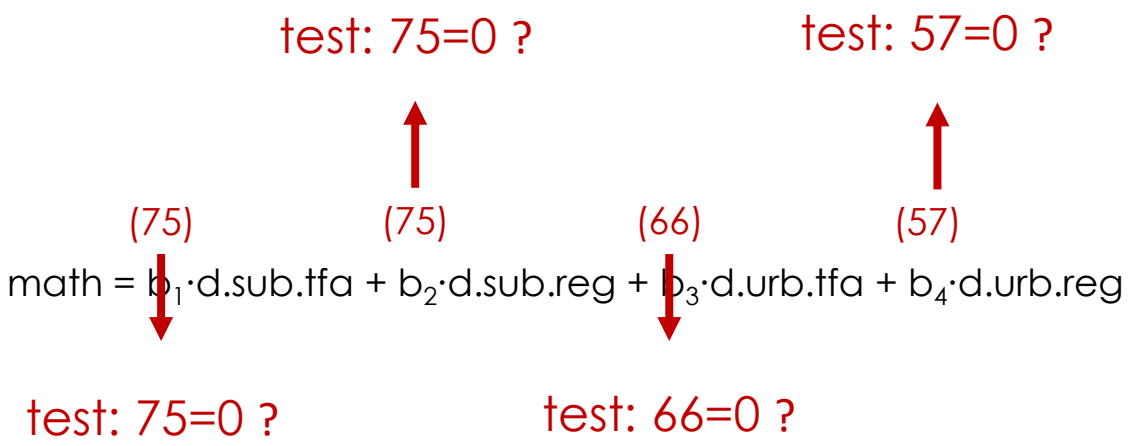
$$math = b_1 \cdot d.sub.tfa + b_2 \cdot d.sub.reg + b_3 \cdot d.urb.tfa + b_4 \cdot d.urb.reg$$

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**



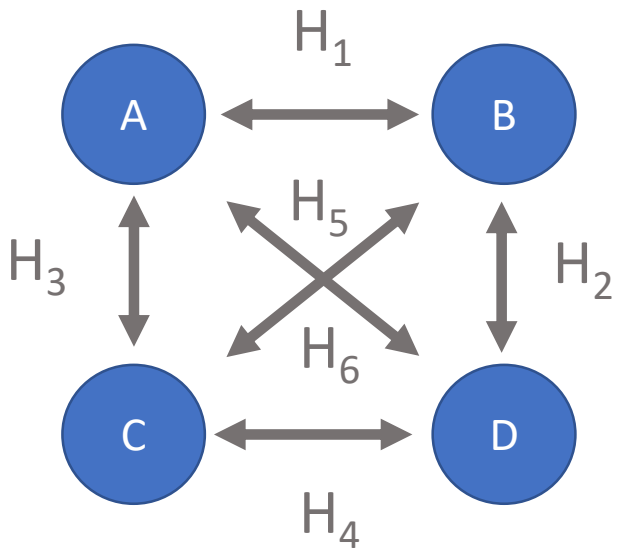
Our research question is whether teachers trained in the Teach for America (TFA) program perform better than teachers trained in the regular program?

None of the tests for significance of coefficients b0 to b3 in the regression reflect meaningful tests.

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)

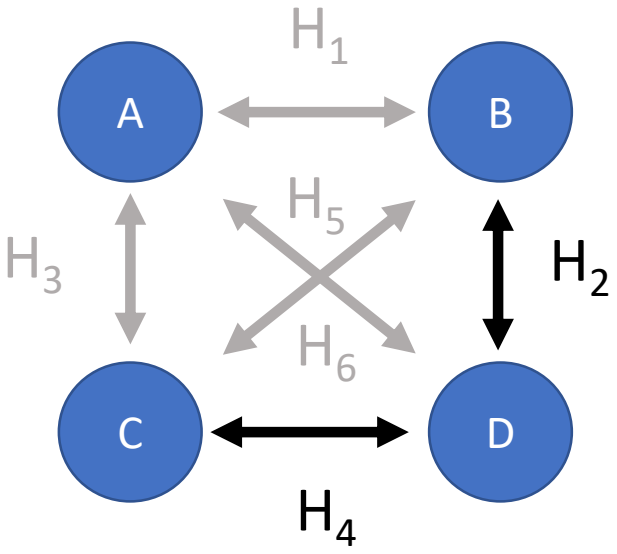
$b_0 = \mathbf{66}$ (urban TFA – **reference group**)

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

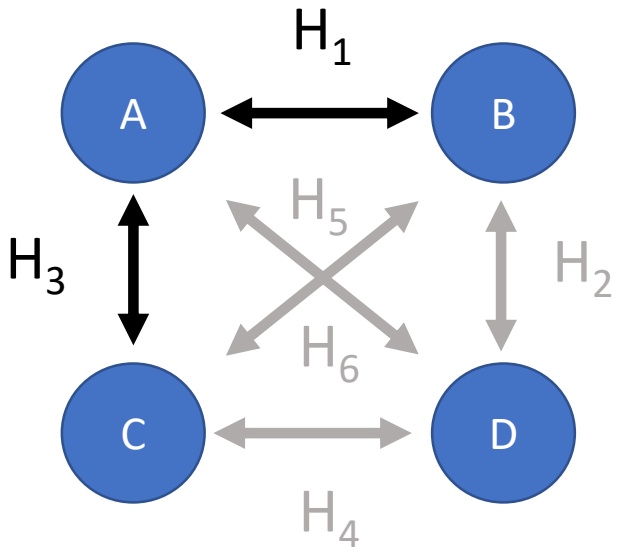
$b_0 + b_2 = 75 - 18 = \mathbf{57}$ (urban regular)

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

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

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

(suburban **x** regular different than suburban + regular ?)

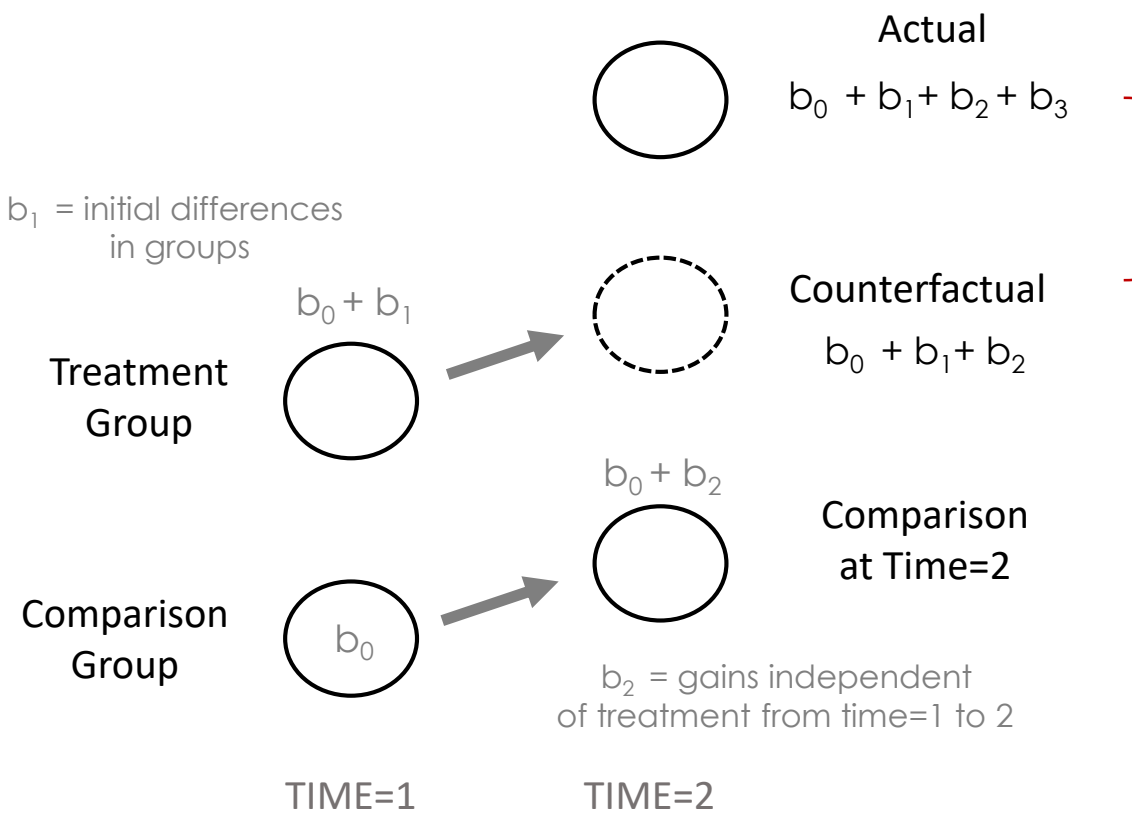


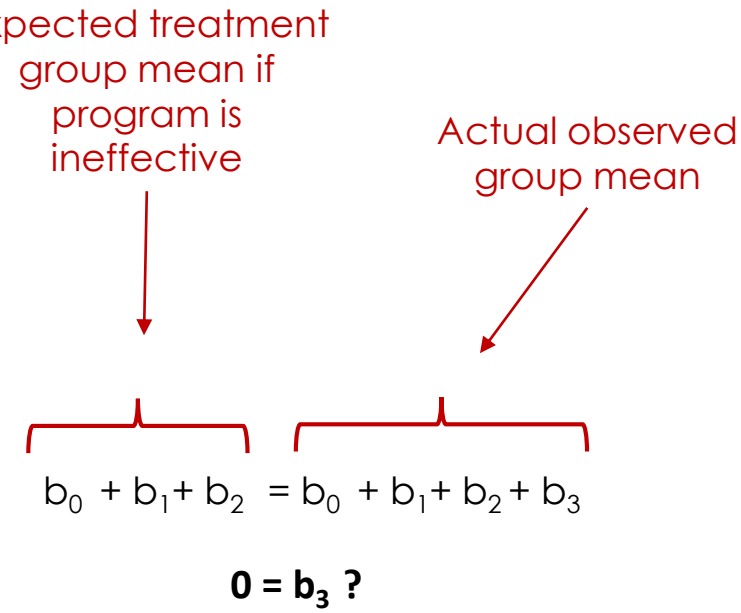
Test for treatment effects in pre-post design with control group:

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