# Regression: groups and interactions

**Topic 14 · Thursday, March 11**

EDUC 15: Statistics for Education Research
UCI School of Education

# Where are we?

## Description

Summarize data

Visualize data

## Prediction

Do these groups differ?

Does one score go with another?

Is the slope real?

## Explanation

Did the program cause it?

**We are here**



Peter the Anteater at a UC Irvine basketball game.

# Plan for today

**ANOVA and t-tests are regression**

**Interactions**

**A sharper test, and the word "caused"**

# ANOVA and t-tests are regression

# A center is a name, not a number

**Regression needs numbers, so we turn each center into yes/no columns**

| A student in | Govindpuri? | Tekhand? |
|---|---|---|
| Chhatarpur | 0 | 0 |
| Govindpuri | 1 | 0 |
| Tekhand | 0 | 1 |

$$\hat{Y} = a + b_1 \times \text{Govindpuri} + b_2 \times \text{Tekhand}$$

**Three groups need two columns: k − 1, which is ANOVA's df between**

The book stops at one predictor. The handout on this topic's page covers two or more

# Each row is a mean or a gap

```
lm(formula = baseline_math ~ center, data = students)
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      11.67081    0.30638  38.093   <2e-16 ***
centerGovindpuri -0.07325    0.52742  -0.139   0.8896
centerTekhand    -0.88889    0.36614  -2.428   0.0155 *
Residual standard error: 3.887 on 616 degrees of freedom
Multiple R-squared:  0.01173, Adjusted R-squared:  0.008524
F-statistic: 3.657 on 2 and 616 DF,  p-value: 0.02638
```

← **a: Chhatarpur's mean**

← **Govindpuri − Chhatarpur**

← **Tekhand − Chhatarpur**

← **$\sqrt{MS_{within}}$**

← **Topic 10's F**

**Start at a and add each gap: 11.67 − 0.07 = 11.60, and 11.67 − 0.89 = 10.78**

**Those are Topic 10's three means. With yes/no columns, every row is about groups**

Data: Muralidharan, Singh, and Ganimian (2019). All 619 students at baseline. Output from R, `summary(lm())`.

# Same split, same F

| In Topic 10 | In regression | SS | df | MS | F |
|---|---|---|---|---|---|
| Between centers | Regression | 110.52 | 2 | 55.26 | **3.66** |
| Within centers | Residual | 9,309.39 | 616 | 15.11 | |
| Total | Total | 9,419.91 | 618 | | |

**Same numbers, to the last decimal: $F(2, 616) = 3.66$, $p = 0.026$**

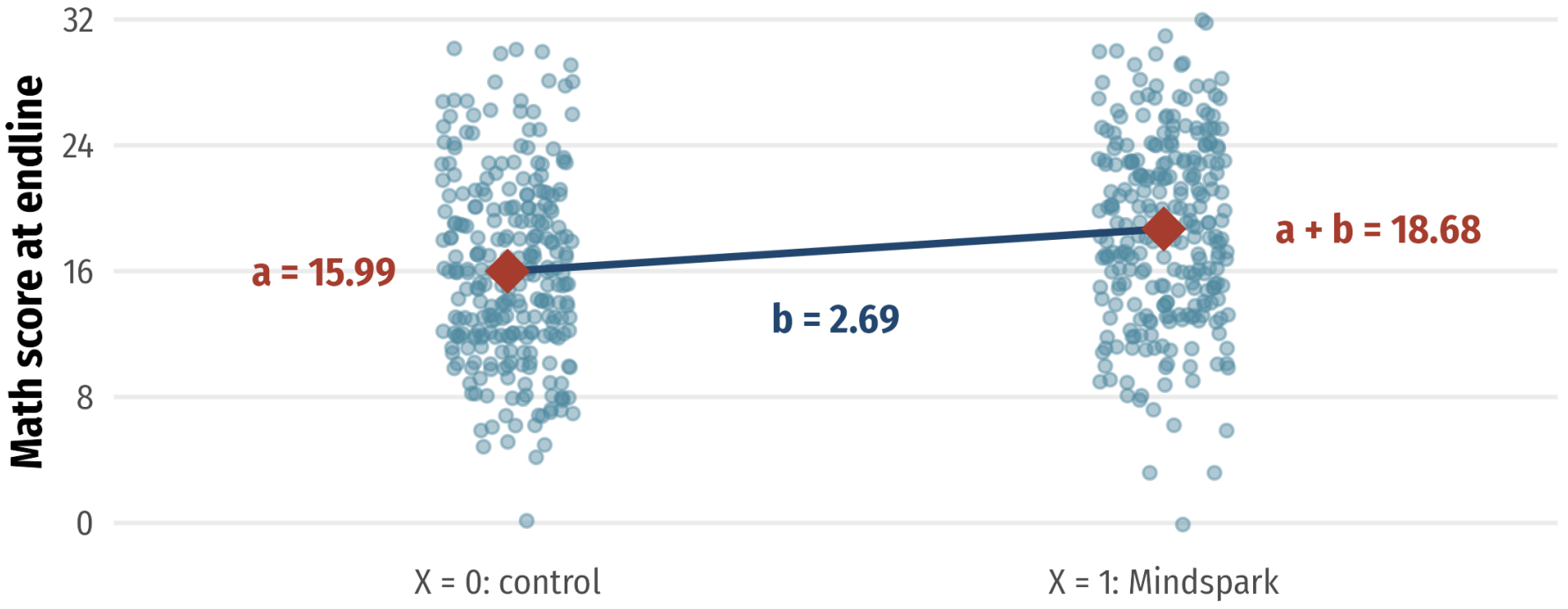
**Topic 10's $\hat{\eta}^2 = 0.012$ is R's Multiple R-squared**

**A one-way ANOVA is a regression on yes/no columns**

With more than one predictor, $r^2$ is written $R^2$

Data: Muralidharan, Singh, and Ganimian (2019).

# With two groups, one column will do




**X can only be 0 or 1, so the line joins two means. Its slope is the gap**

Data: Muralidharan, Singh, and Ganimian (2019). 539 students. Dots are nudged so equal scores don't hide each other.

# The slope is Topic 9's gap

```
lm(formula = endline_math ~ treat, data = mindspark)
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  15.9927     0.3520  45.430  < 2e-16 ***
treat         2.6915     0.5011   5.371 1.17e-07 ***
Residual standard error: 5.816 on 537 degrees of freedom
Multiple R-squared:  0.05098, Adjusted R-squared:  0.04922
F-statistic: 28.85 on 1 and 537 DF,  p-value: 1.167e-07
```

← **a: the control mean**
← **b: the gap, and its σ̂**
← **√pooled variance**
← **$r^2$**
← **$F = t^2$**

**$\hat{Y} = 15.99 + 2.69X$. At $X = 0$, $\hat{Y} = 15.99$**

**At $X = 1$, $\hat{Y} = 15.99 + 2.69 = 18.68$**

**The Std. Error, 0.501, is Topic 9's $\hat{\sigma}_{M_1-M_2}$. The t, 5.37, is Topic 9's t**

Data: Muralidharan, Singh, and Ganimian (2019). Output from R, `summary(lm())`.

# One gap, four tests, one answer

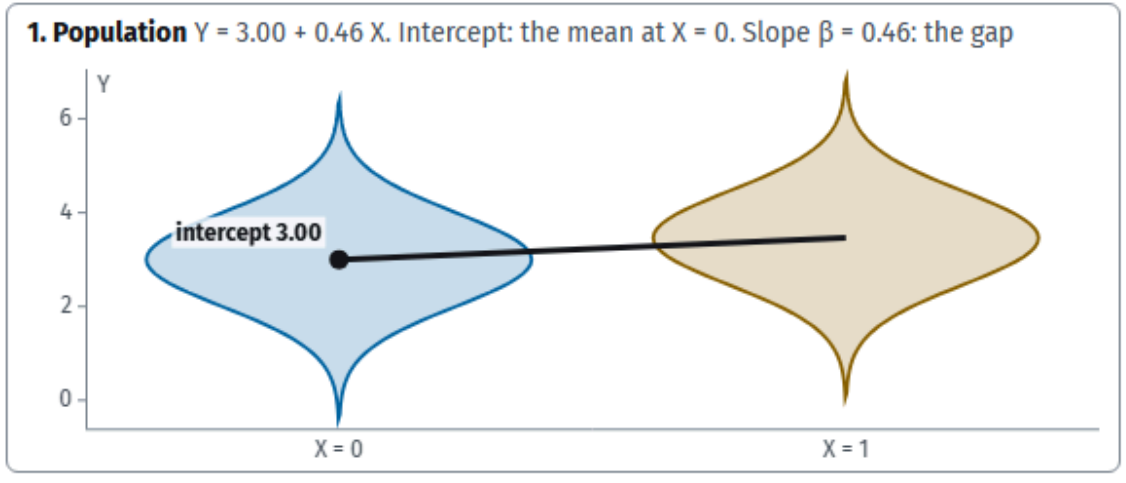
| Test of the gap | Where we met it | Mindspark at endline |
|---|---|---|
| Two-sample t | Topic 9 | t(537) = 5.37 |
| One-way ANOVA | Topic 10 | F(1, 537) = 28.85 |
| Correlation with a yes/no X | Topic 11 | r = 0.23, t(537) = 5.37 |
| **Regression slope** | **Today** | **b = 2.69, t(537) = 5.37** |

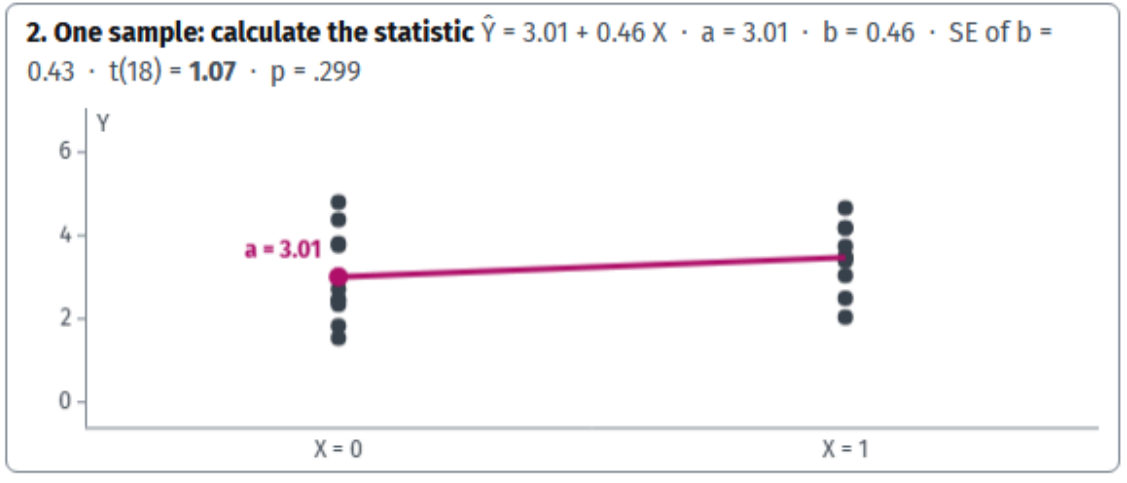
**$5.37^2 = 28.85$. And $r^2 = 0.051$ is the share of variability between the groups**

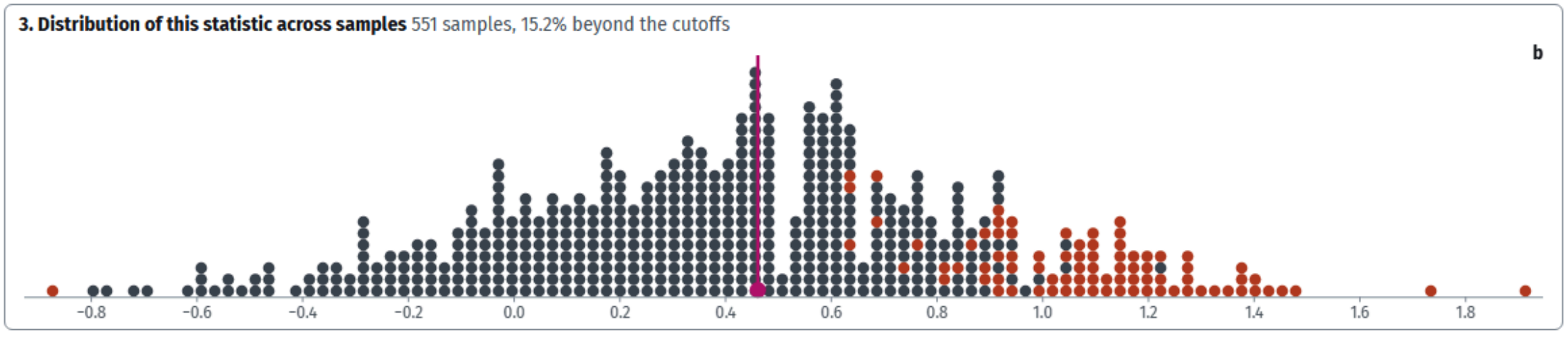
**The book teaches these one at a time. They're one model: regression**

Data: Muralidharan, Singh, and Ganimian (2019).

# Step 7b: a slope for a yes/no X

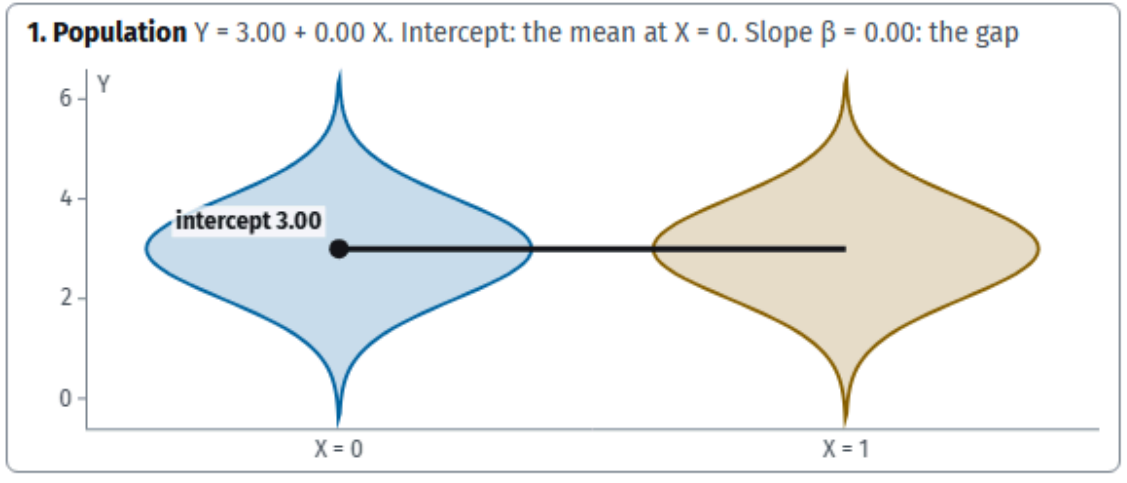


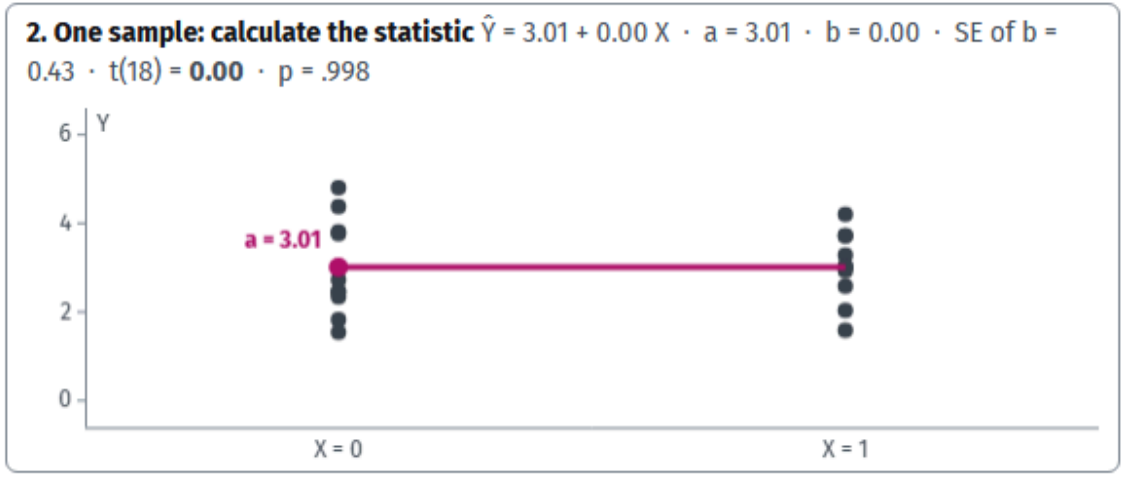



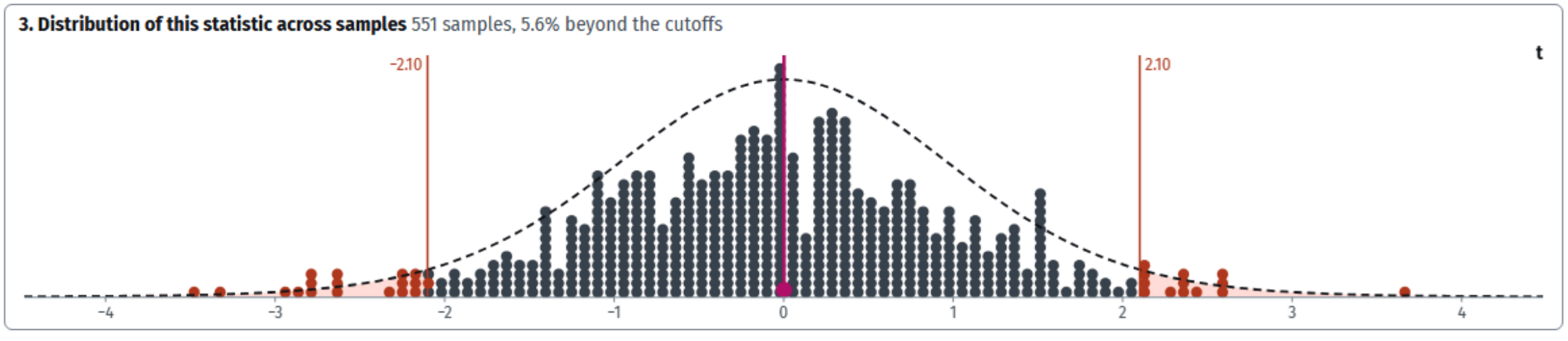




Simulator after Kristoffer Magnusson, rpsychologist.com. On your phone: [the simulator page](#).

# Under $H_0$, it's step 4's t curve




Simulator after Kristoffer Magnusson, rpsychologist.com.

# Try it: One center's output

```
lm(formula = endline_math ~ treat, data = chhatarpur)
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  18.3514     0.6836   26.85   <2e-16 ***
treat         1.6032     0.9956    1.61     0.11
Residual standard error: 5.88 on 138 degrees of freedom
Multiple R-squared:  0.01844, Adjusted R-squared:  0.01133
```

**These are Chhatarpur's 140 students, with treat = 0 or 1. Find the control mean, the Mindspark mean and F**

**Then decide with α = 0.05. Work alone, compare with a neighbor, then vote with hands: reject or maintain?**

Data: Muralidharan, Singh, and Ganimian (2019). Output from R, `summary(lm())`.

# Try it: Answer

**Control mean: a = 18.35**

**Mindspark mean: a + b = 19.95**

**$F = t^2 = 1.61^2 = 2.59$**

**p = 0.110 is above 0.05, so we maintain $H_0$: t(138) = 1.61**

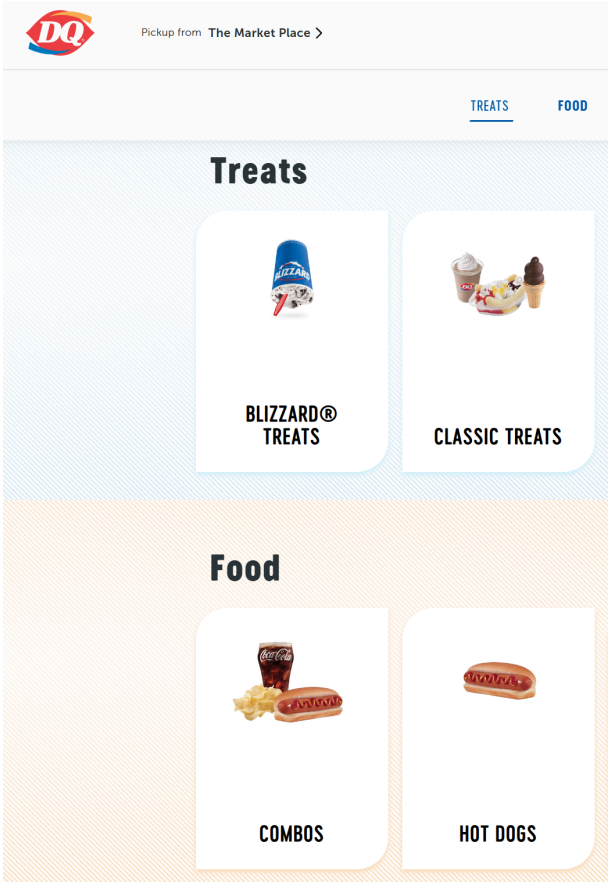
**With only 140 students, this gap's standard error is 1.00, not 0.50**

**Maintaining $H_0$ doesn't mean no effect here. And a gap of 1.60 isn't 2.69. Could the effect depend on who you are?**

Data: Muralidharan, Singh, and Ganimian (2019).

# Interactions

# Dairy Queen sells hot dogs







Menu and sign: Dairy Queen.

# Does chili cheese cost extra?

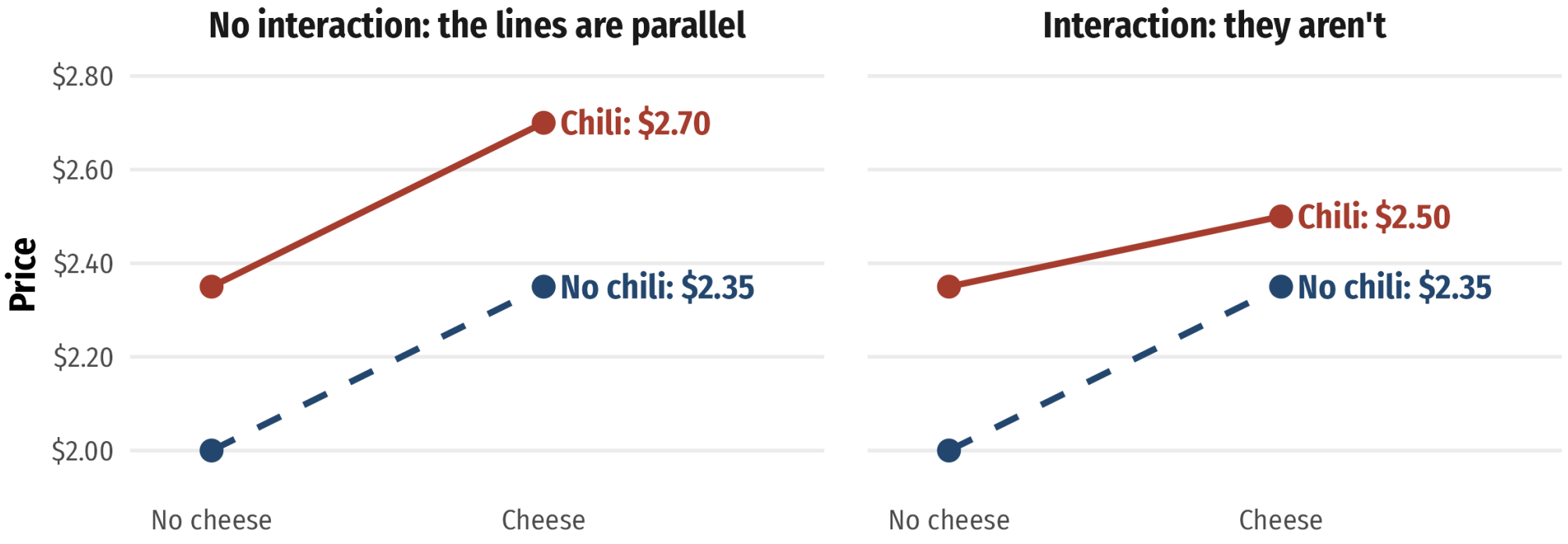
| | No cheese | Cheese |
|---|---|---|
| No chili | $2.00 | $2.35 |
| Chili | $2.35 | $2.70 |

Chili alone adds $0.35. Cheese alone adds $0.35

Both together add $0.70: the sum of the two

An **interaction** is what a combination adds beyond the sum of its parts. Here, it's $0

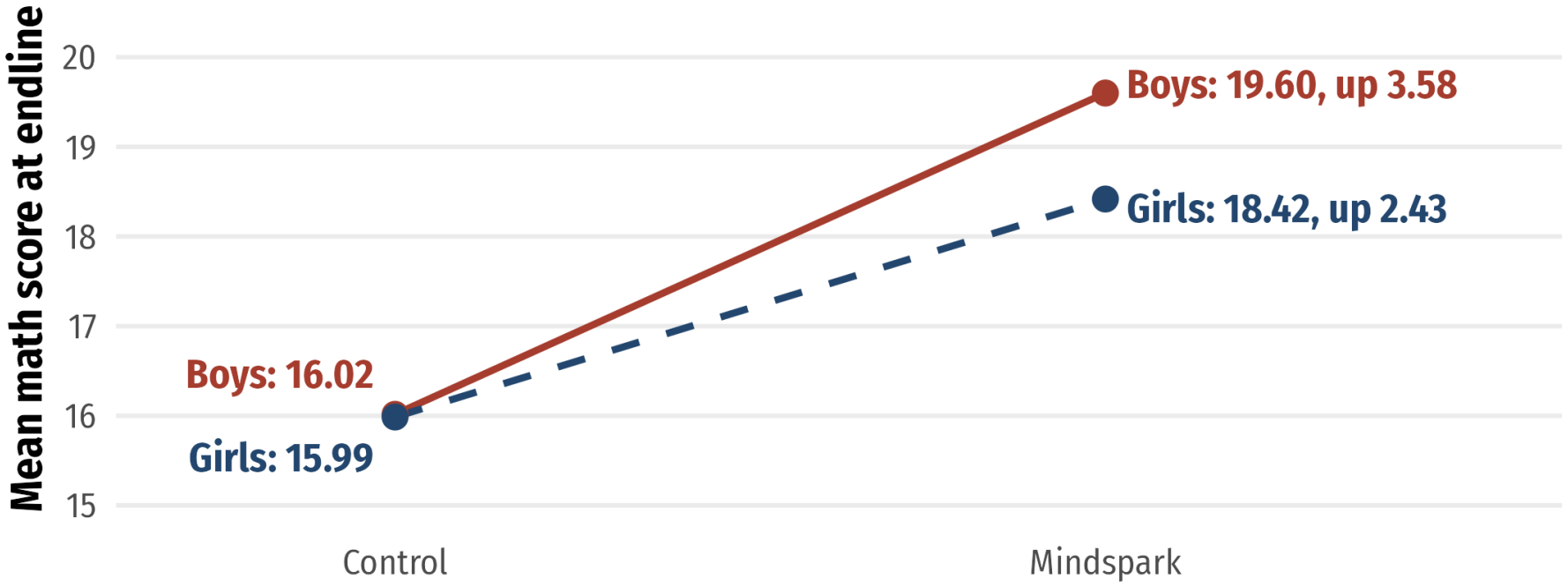
# Parallel lines mean no interaction




**On the right, what chili adds depends on whether you have cheese**

The right-hand price of $2.50 is made up.

# Mindspark's lines aren't quite parallel




**The gap is 3.58 for boys and 2.43 for girls. Is that 1.15-point difference real?**

Data: Muralidharan, Singh, and Ganimian (2019). Boys: 65 control, 60 Mindspark. Girls: 208 control, 206 Mindspark.

# An interaction is one more column

| Group | treat | female | treat × female | n |
|---|---|---|---|---|
| Control boys | 0 | 0 | 0 | 65 |
| Mindspark boys | 1 | 0 | 0 | 60 |
| Control girls | 0 | 1 | 0 | 208 |
| Mindspark girls | 1 | 1 | **1** | 206 |

$$\hat{Y} = a + b_1 \,\text{treat} + b_2 \,\text{female} + b_3 \,(\text{treat} \times \text{female})$$

**The product is 1 for one group only. So $b_3$ is the extra for that group**

Data: Muralidharan, Singh, and Ganimian (2019).

# R writes the product as treat:female

```
lm(formula = endline_math ~ treat * female, data = mindspark)
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  16.01538    0.72150  22.197  < 2e-16 ***
treat         3.58462    1.04139   3.442 0.000622 ***
female       -0.02981    0.82658  -0.036 0.971247
treat:female -1.15272    1.18803  -0.970 0.332349
Residual standard error: 5.817 on 535 degrees of freedom
Multiple R-squared:  0.05438, Adjusted R-squared:  0.04908
F-statistic: 10.26 on 3 and 535 DF,  p-value: 1.419e-06
```

← **a**

← $\mathbf{b_1}$

← $\mathbf{b_2}$

← $\mathbf{b_3}$

$$\hat{Y} = 16.02 + 3.585\ \text{treat} - 0.03\ \text{female} - 1.153\ \text{treat}\times\text{female}$$

**Read the four coefficients as groups. Don't start with the stars**

Data: Muralidharan, Singh, and Ganimian (2019). Output from R, `summary(lm())`.

# Each coefficient compares two groups

| | Value | What it says |
|---|---|---|
| a | 16.02 | The mean of control boys, where every column is 0 |
| $b_1$: treat | 3.585 | Mindspark's effect **for boys**, where female = 0 |
| $b_2$: female | −0.03 | Girls minus boys **in the control group** |
| $b_3$: treat × female | −1.153 | How much smaller the effect is for girls |

**Mindspark girls: 16.02 + 3.585 − 0.03 − 1.153 = 18.42. Every column is 1, so add all four**

**With a product in the model, each b holds where the other column is 0**

You'll also read: moderation, or heterogeneous, conditional or differential effects

Data: Muralidharan, Singh, and Ganimian (2019).

# The four steps: is the interaction real?

| Step | In words | Our numbers |
|---|---|---|
| 1. State $H_0$ | Same effect for girls and boys | $H_0$: $\beta_3 = 0$ |
| 2. Set the criteria | Two tails, $\alpha = 0.05$ | t(535) beyond ±1.964 |
| 3. Compute t | Estimate / Std. Error | −1.153 / 1.188 = −0.97 |
| 4. Decide | Inside the cutoffs | **Maintain $H_0$**, p = 0.332 |

**We can't say the effect differs by gender. We can't say it's the same either: the 95% interval for $\beta_3$ is [−3.49, 1.18]**

**$b_3$'s standard error is 1.19, over twice the 0.50 for the overall gap. Only 125 students are boys, and that limits this test**

Data: Muralidharan, Singh, and Ganimian (2019).

# Play it: Mindspark and math tuition

```
lm(formula = endline_math ~ treat * tuition, data = mindspark)
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    15.5057     0.4348  35.658  < 2e-16 ***
treat           2.9576     0.6141   4.816 1.91e-06 ***
tuition         1.3773     0.7370   1.869   0.0622 .
treat:tuition  -0.5034     1.0574  -0.476   0.6342
```

**100** The control mean without tuition

**200** What Mindspark adds without tuition

**300** What Mindspark adds with tuition

**400** Does the effect depend on tuition? α = 0.05

**tuition = 1 means extra math tuition. Answer with a neighbor, as a question**

Data: Muralidharan, Singh, and Ganimian (2019). Output from R, `summary(lm())`.

# Play it: Answers

**100: What is 15.51?**
a: every column is 0

**200: What is 2.96?**
$b_1$: the effect where tuition = 0

**300: What is 2.45?**
$b_1 + b_3$: 2.96 − 0.50

**400: What is maintain $H_0$?**
t(529) = −0.48, p = 0.634

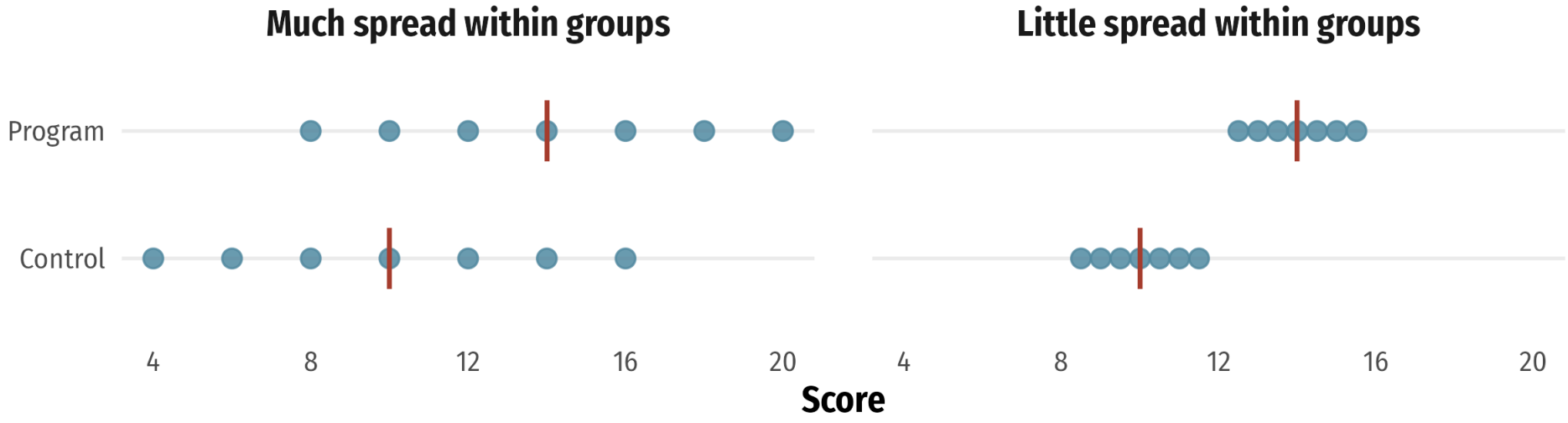
**The lines are close to parallel. Mindspark helped about as much either way**

**The lottery set Mindspark, but not tuition. So $b_2$ = 1.38 is a difference, not an effect**

Data: Muralidharan, Singh, and Ganimian (2019). 533 students with a tuition answer.

# A sharper test, and the word "caused"

# Spread hides a gap




**Same gap, less spread: a smaller standard error and a bigger t**

**That's more power, so less risk of a Type 2 error. α stays at 0.05**

After a figure in Gravetter et al., *Essentials of Statistics for the Behavioral Sciences*, 10e. Scores made up.

# Predict it: Which one sharpens the test?

Respond at **PollEv.com/adb**

**With only treat in the model, the gap's standard error is 0.501. We add one more column. Which makes that standard error smallest?**

**A** Gender

**B** Baseline math

**C** Both about the same

**D** Neither: the lottery already took care of it

Data: Muralidharan, Singh, and Ganimian (2019).

# A good covariate shrinks the misses

| Model | b for treat | Std. Error | t | Typical miss |
|---|---|---|---|---|
| treat alone | 2.69 | 0.501 | 5.37 | 5.82 |
| treat + gender | 2.70 | 0.501 | 5.39 | 5.82 |
| **treat + baseline** | 3.00 | 0.427 | 7.04 | 4.94 |
| Gain as the outcome (Topic 9) | 3.09 | 0.432 | 7.15 | 5.01 |

**Baseline takes the typical miss from 5.82 to 4.94 points. Gender leaves it alone**

**b moves too, because the lottery left the control group 0.40 ahead by chance. Gains assume baseline's slope is 1. Regression estimates it: 0.78**

**A covariate is a predictor added to sharpen a test, by explaining Y**

Data: Muralidharan, Singh, and Ganimian (2019). Typical miss is R's Residual standard error.

# The lottery earns the word "caused"

We can't watch the same student both with and without Mindspark

A lottery sorts students by chance. So motivation, family and past scores balance out on average, even the things nobody measured

Apart from chance, the program is then the only thing that differs between the groups

**And our test deals with chance: t(537) = 5.37. What's left is the program**

Data: Muralidharan, Singh, and Ganimian (2019).

# Same math, different sentence

| X | R prints | Who set X? | We may say |
|---|---|---|---|
| Tekhand | b = −0.89, t = −2.43 | Not a lottery | Lower than Chhatarpur |
| Baseline | b = 0.76, t = 13.46 | Past learning | Goes with endline |
| **treat** | **b = 2.69, t = 5.37** | **A lottery** | **Mindspark raised scores** |

**R prints all three the same way. It doesn't know how the data were made**

**The test says a gap is real. Only the design says why**

Data: Muralidharan, Singh, and Ganimian (2019).

# Back to January: real, or just luck?

| Question | What we found | Topic |
|---|---|---|
| Is the gap real? | Yes: t(537) = 5.37, p < 0.001 | 9 |
| How big is it? | 2.69 points, g = 0.46, 95% CI [1.71, 3.68] | 9 |
| Isn't it just growth? | Everyone rose: control 4.74, Mindspark 7.82 | 9 |
| Does it depend on gender? | We can't tell: t(535) = −0.97 | 14 |
| Can we sharpen it? | With baseline: 3.00 points, t = 7.04 | 14 |
| Did Mindspark cause it? | Yes: a lottery set who got it | 14 |

**That's a full research story: describe, test, size it, rule out rivals, and say why**

Data: Muralidharan, Singh, and Ganimian (2019).

# The blueprint, finished

| Blueprint step | Statistic | Bottom of the fraction | df |
|---|---|---|---|
| 1. One score | z | $\sigma$ | None |
| 2. One mean, σ known | z | $\sigma_M$ | None |
| 3a. One mean, σ unknown | t | $\hat{\sigma}_M$ | n − 1 |
| 3b. Mean gain | t | $\hat{\sigma}_M$ of the gains | n − 1 |
| 4. Two means | t | $\hat{\sigma}_{M_1-M_2}$ | $n_1 + n_2 - 2$ |
| 5. Two or more means | F | $MS_{within}$ | k − 1 and N − k |
| 6. A correlation | t | $\hat{\sigma}_r$ | n − 2 |
| 7a. A slope | t | $\hat{\sigma}_b$ | n − 2 |
| **7b. A slope for a yes/no X** | **t** | **$\hat{\sigma}_b$, which is $\hat{\sigma}_{M_1-M_2}$** | **n − 2** |

**z, t, F, r, b: one idea. What we saw, over what chance alone would give**

# Where are we?

## Description

Summarize data

Visualize data

## Prediction

Do these groups differ?

Does one score go with another?

Is the slope real?

## Explanation

Did the program cause it?

A lottery lets us answer yes

**All three, done**

# What you can do now

Read a regression on yes/no columns as group means and gaps

Show that ANOVA and the two-sample t are the same regression model

Spot an interaction in a plot, and say what each coefficient means

Say why a good covariate sharpens a test, and a weak one doesn't

Say why a lottery lets us say "caused", and when we can't

# Next steps

**Friday, March 12: Lab 10**

Quiz: final review. Bring a calculator and your questions

**Finals week, March 13 to 19**

The final exam. Date to come

**Nothing more is due in MindTap**