ANOVA

Topic 10 · Thursday, February 25

EDUC 15: Statistics for Education Research
UCI School of Education

Where are we?

Description

Summarize data

Visualize data

Prediction

Do these groups differ in the population?

We are here



Explanation

Peter the Anteater at a UC Irvine basketball game.

Same four steps, more means

Step 1: State a hypothesis about the means of two or more populations

Step 2: Set the criteria: how large must F be to count as unexpected?

Step 3: Observe a sample from each population. How unexpected is the spread between the group means, given the spread within the groups?

Step 4: Reject or maintain H_0

Plan for today

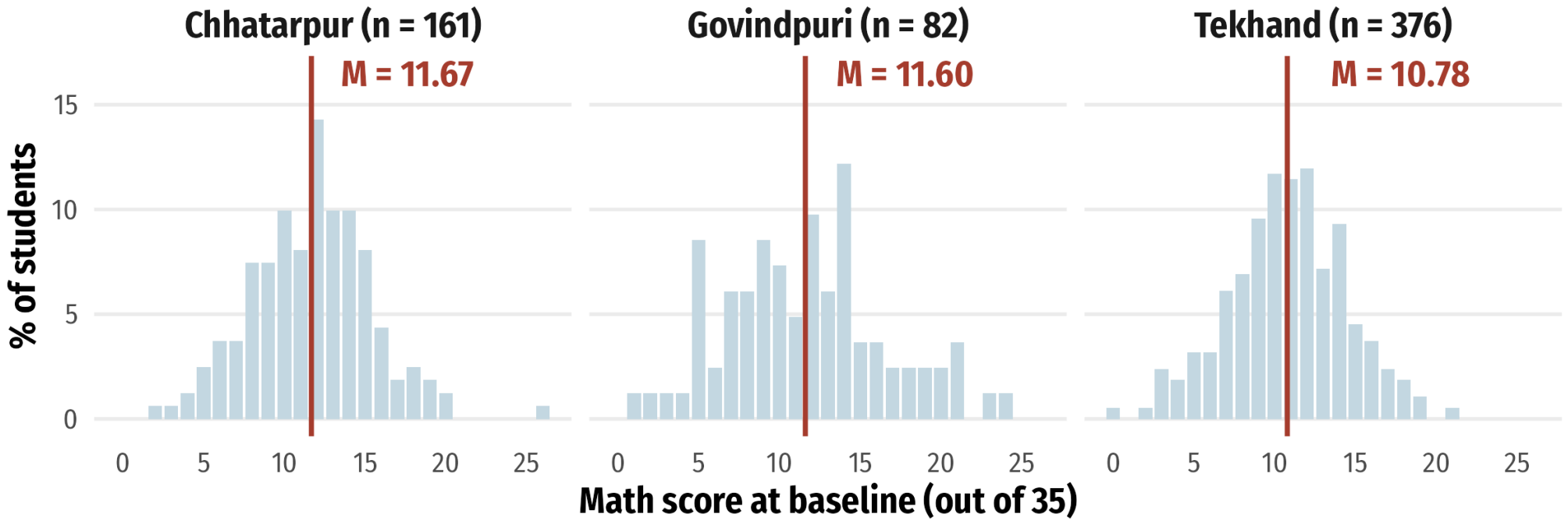
Why not more t-tests?

Splitting the variability

The test, and what comes after

Why not more t-tests?

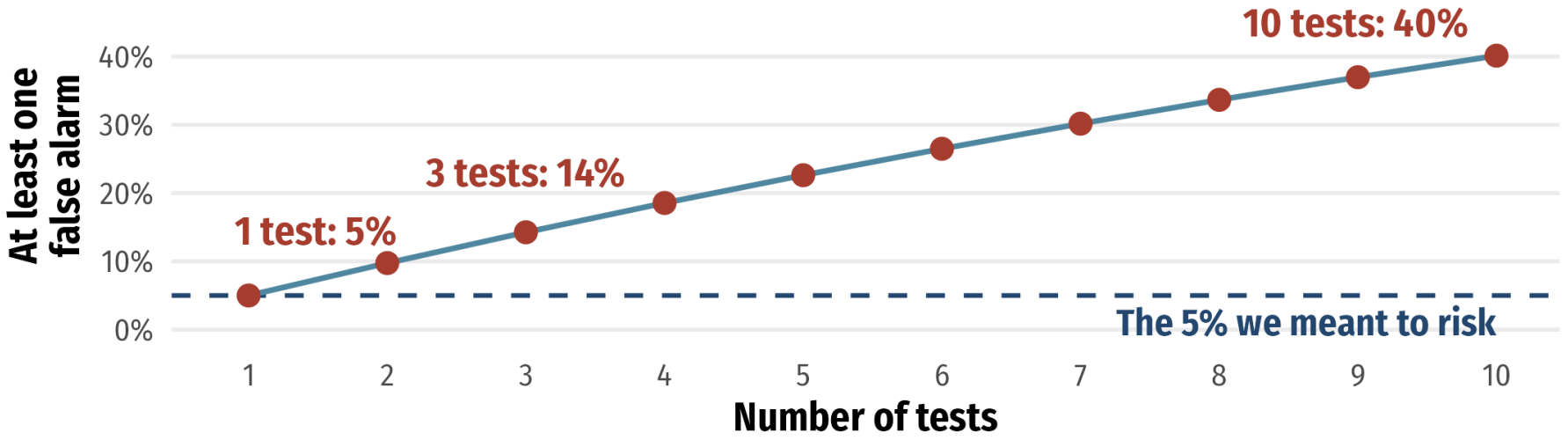
Three centers, three means



Did the centers differ before the program, or is this sampling error?

Data: Muralidharan, Singh, and Ganimian (2019). 619 students at baseline.

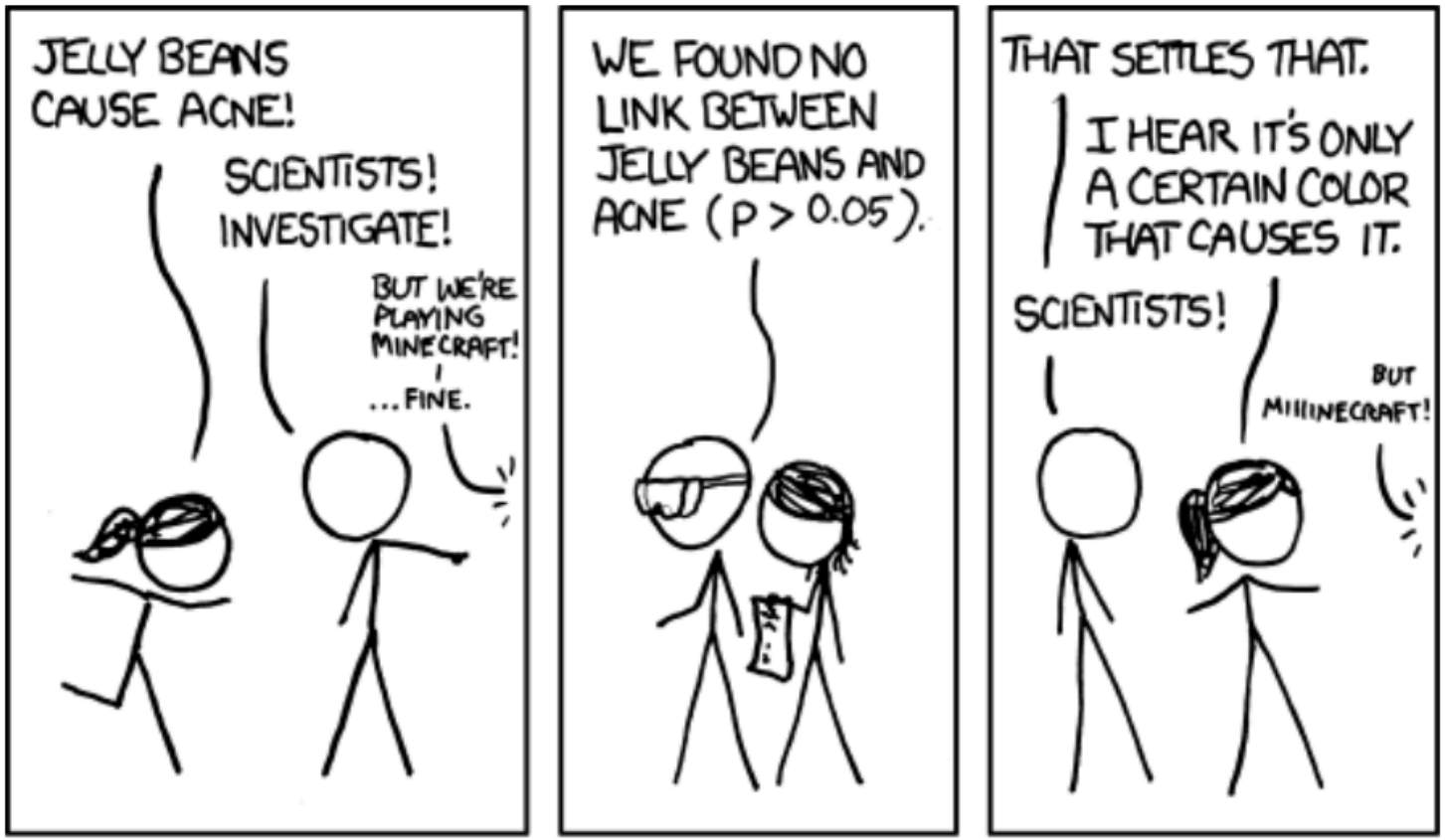
More tests, more false alarms



$$\text{Chance of at least one false alarm} = 1 - (1 - \alpha)^{\text{tests}}$$

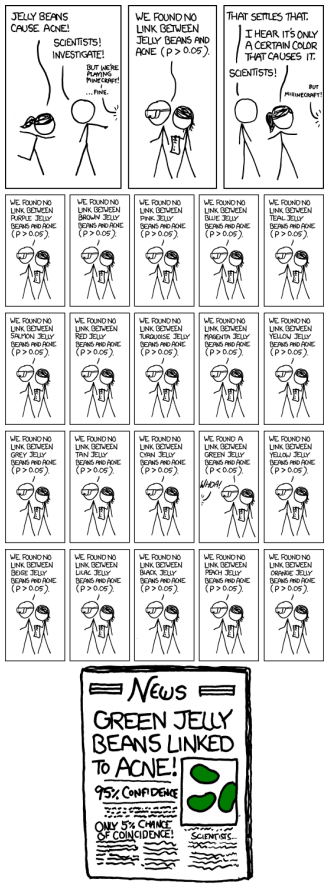
Three centers need 3 t-tests, and five groups need 10

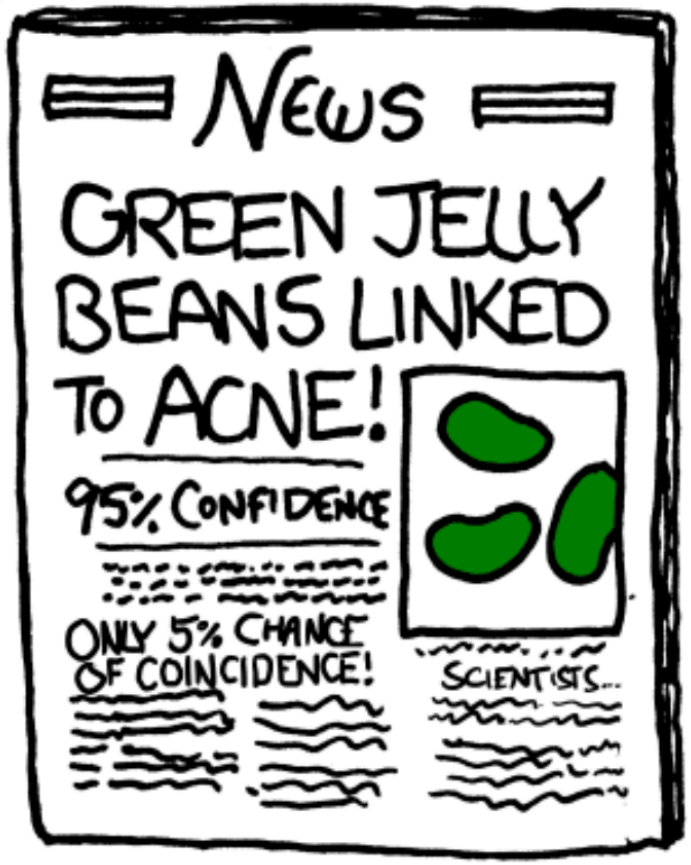
Green jelly beans



Randall Munroe, "Significant," xkcd 882, CC BY-NC 2.5.

Test 20 colors, and one will go off





Randall Munroe, "Significant," xkcd 882, CC BY-NC 2.5.

One question about all the means

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

H_0: all population means are equal

H_1: at least one differs from another

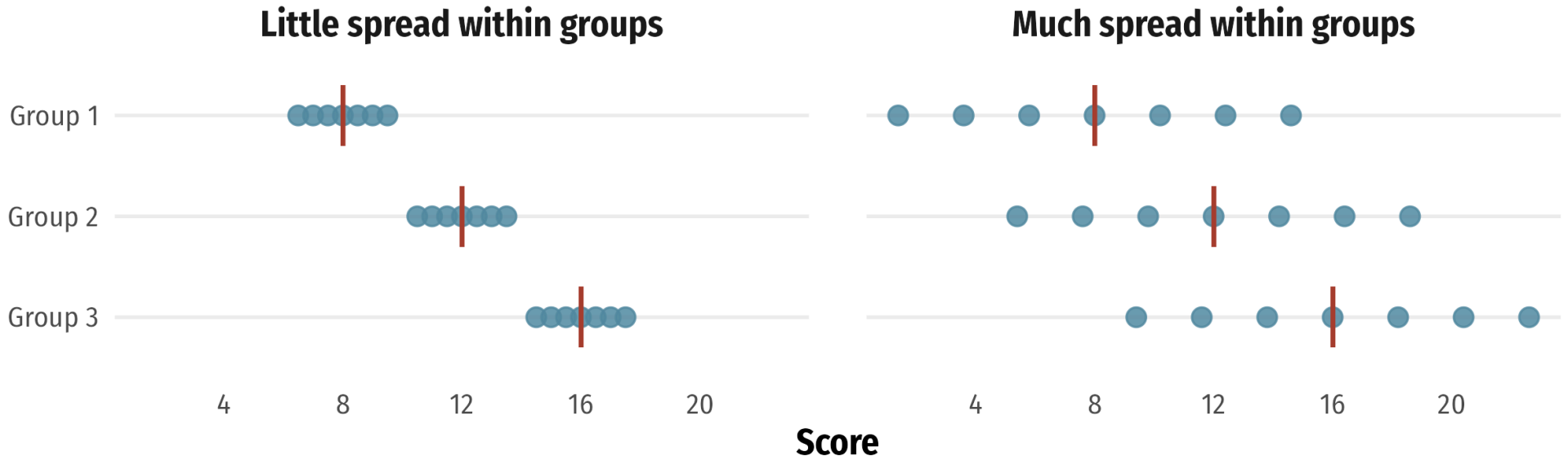
One test, so α stays at 0.05 for the whole question

Factor: the variable that makes the groups. Here, center

Levels: the groups of that factor. Here, 3 centers

ANOVA stands for analysis of variance

Same gaps, but the spread matters



On the left, the gaps stand out. On the right, they could be chance

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

F compares two variances

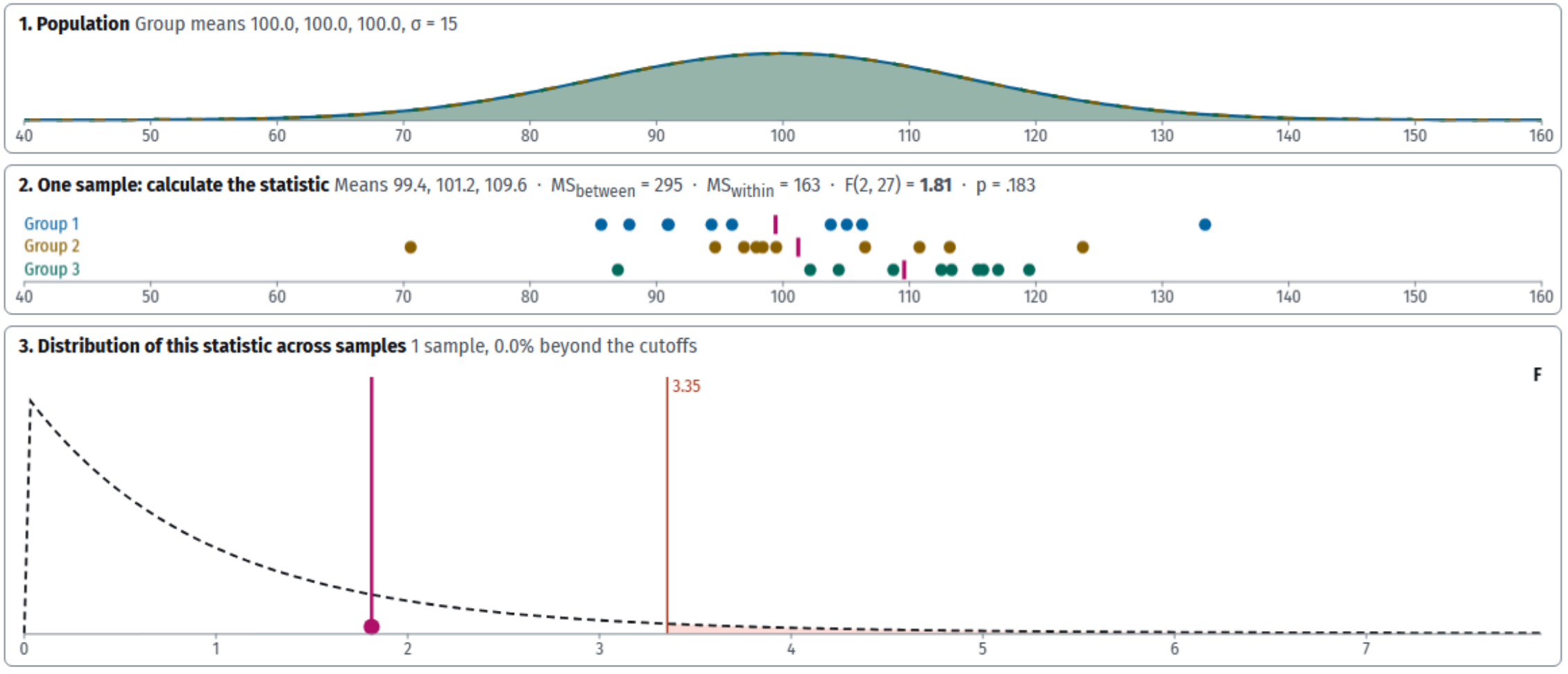
$$F = \frac{\text{variance between the group means}}{\text{variance within the groups}} = \frac{MS_{\text{between}}}{MS_{\text{within}}}$$

Top: how far the group means sit apart. It holds any real difference, plus chance

Bottom: how far people sit from their own group's mean. It holds chance only

MS means mean square: ANOVA's word for a variance

Step 5: three populations



Simulator after Kristoffer Magnusson, rpsychologist.com.

Predict it: No real differences

Respond at **PollEv.com/adb**

All three population means are the same, so H_0 is true. Draw three samples, compute F, and repeat thousands of times. What does F average out to?

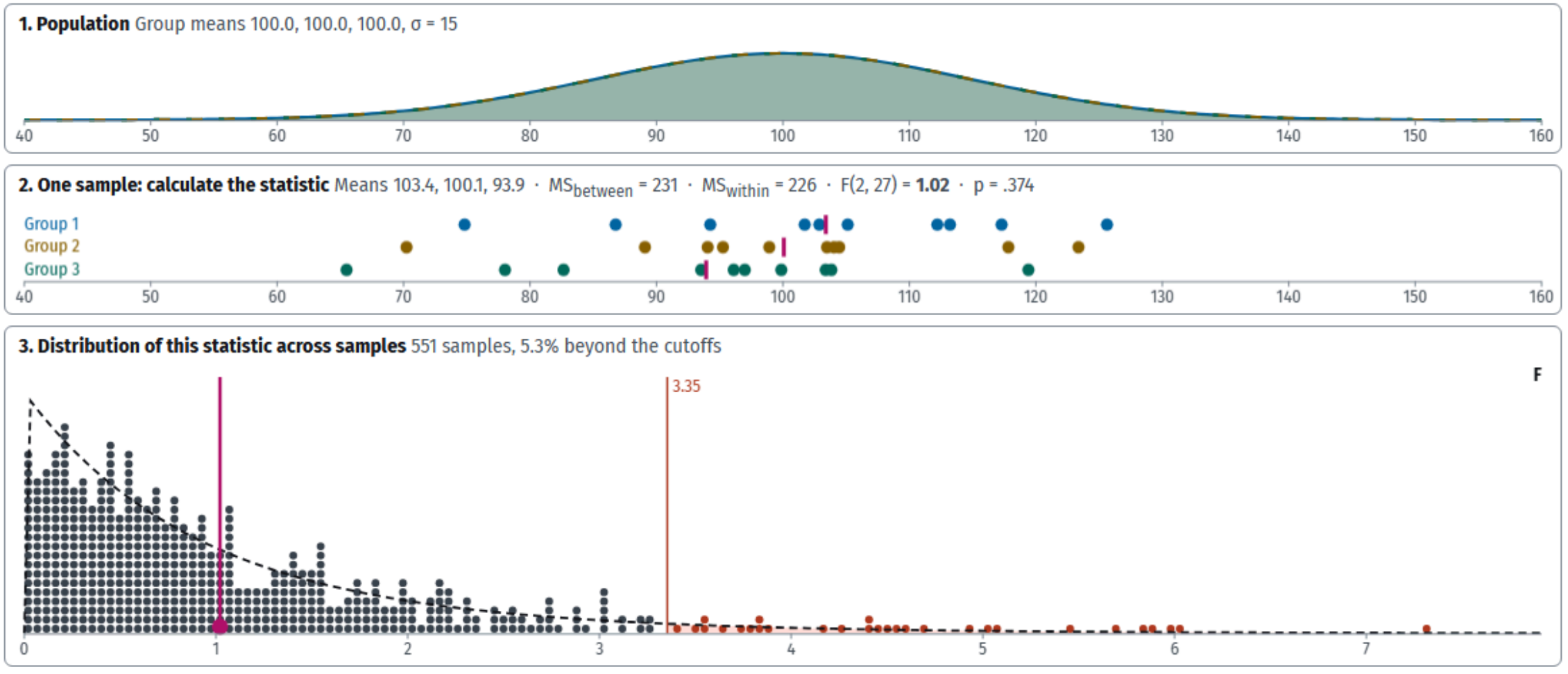
A About 0

B About 1

C About 3

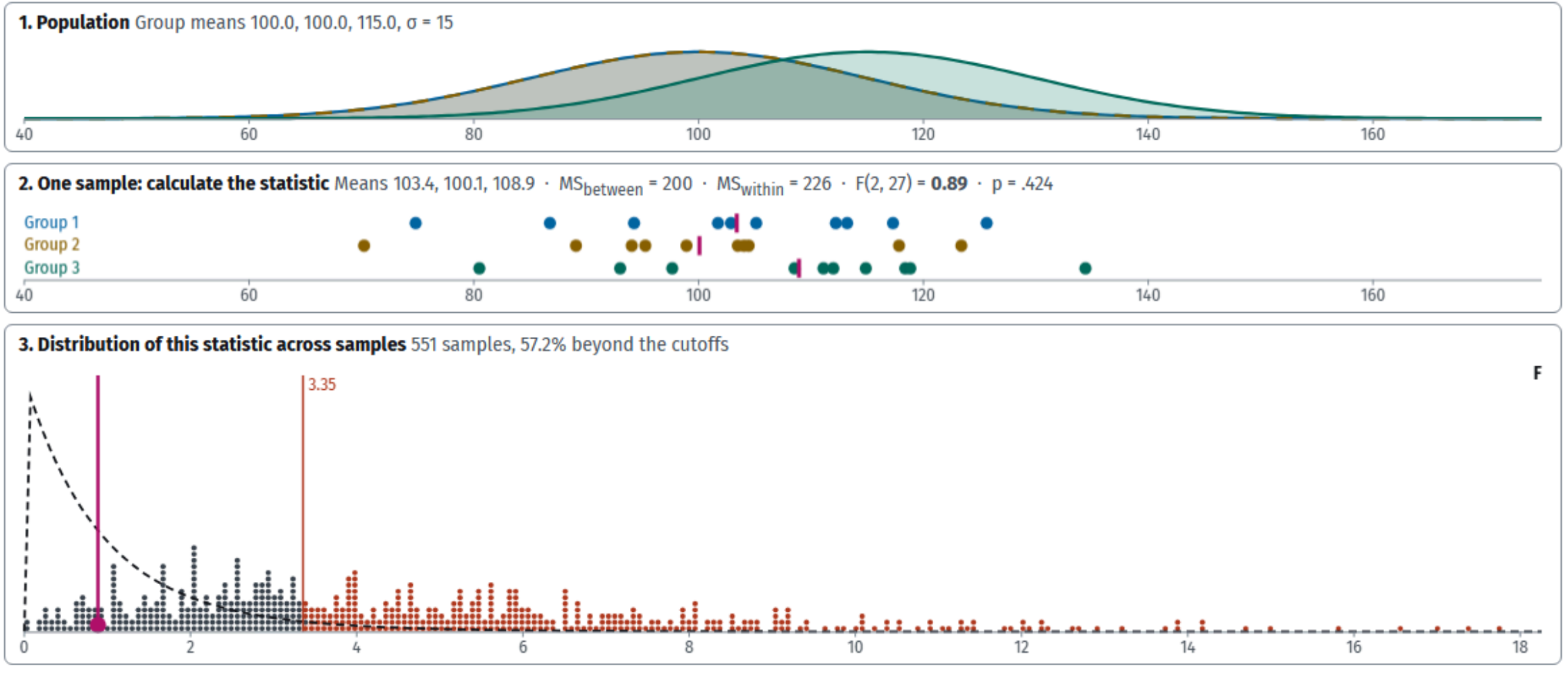
D It depends on the size of σ

Under H_0, F averages about 1



Simulator after Kristoffer Magnusson, rpsychologist.com.

When means differ, the pile moves



Simulator after Kristoffer Magnusson, rpsychologist.com.

MS between grows when means differ

Population means	MS within averages	MS between averages	F averages
100, 100, 100 (H_0 is true)	225	224	1.08
100, 100, 115 (H_1 is true)	225	975	4.70

Here $\sigma^2 = 225$. MS within estimates σ^2 in both worlds

MS between estimates σ^2 only when H_0 is true. If the means differ, it adds their real spread

So F near 1 fits H_0, and a large F is evidence against it

Splitting the variability

The total splits into two parts

Total variability

↙ ↘

Between the groups

Any real differences, plus chance

Goes on top of F

Within the groups

Chance only

Goes on the bottom of F

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

First, each center's own spread

Center	n	Sum of scores T	M	SS within the center
Chhatarpur	161	1,879	11.67	2,229.6
Govindpuri	82	951	11.60	2,039.7
Tekhand	376	4,054	10.78	5,040.1
All three	619	6,884	11.12	**9,309.4**

SS_{within} adds up each group's SS: every score against its **own** group's mean

SS_{total} treats everyone as one group, against the overall mean. Here it's 9,419.9

In the book: k = 3 groups, N = 619 students, and G = 6,884 is the grand total of T

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

Next, the means' own spread

Center	n	M	M − overall mean	n × (M − overall mean)²
Chhatarpur	161	11.67	0.55	48.6
Govindpuri	82	11.60	0.48	18.6
Tekhand	376	10.78	−0.34	43.3
Add them up				**110.5**

Every student counts, so a bigger center gets more weight. The overall mean is 11.12

Check the split: 9,309.4 + 110.5 = 9,419.9. The parts add up to SS_{total}

In the book: $SS_{between} = \Sigma T^2 / n - G^2 / N$. It's the same number

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

df split the same way

Source	df	Our numbers	Why
Between	$k - 1$	$3 - 1 = 2$	3 means, and 1 df spent on the overall mean
Within	$N - k$	$619 - 3 = 616$	Each group spends 1 df on its own mean
Total	$N - 1$	$619 - 1 = 618$	Everyone, around 1 overall mean

Check the split: $2 + 616 = 618$. F has two df, and we report both

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

A mean square is a variance

$$MS_{\text{between}} = SS_{\text{between}} \,/\, df_{\text{between}} = 110.52 \,/\, 2 = 55.26$$

$$MS_{\text{within}} = SS_{\text{within}} \,/\, df_{\text{within}} = 9309.39 \,/\, 616 = 15.11$$

$$F = MS_{\text{between}} \,/\, MS_{\text{within}} = 55.26 \,/\, 15.11 = 3.66$$

MS_{within} is the pooled variance from Topic 9, now with three groups

In the book: each MS is also written s^2. Both are estimates, like $\hat{\sigma}^2$

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

The source table holds all of it

Source	SS	df	MS	F
Between centers	110.52	2	55.26	**3.66**
Within centers	9,309.39	616	15.11	
Total	9,419.91	618		

Read it left to right: SS / df = MS, then MS between / MS within = F

Software prints this same table. Now you can read it

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

Try it: Centers at endline

Source	SS	df	MS	F
Between centers	?	?	?	?
Within centers	18,357.4	?	?	
Total	19,143.5	?		

3 centers, 539 students. The cutoff is F(2, 536) = 3.01. Fill in the table, then decide

Work alone for two minutes. Compare with a neighbor. Then we vote: reject or maintain?

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

Try it: Answer

Source	SS	df	MS	F
Between centers	**786.0**	**2**	**393.0**	**11.48**
Within centers	18,357.4	**536**	**34.25**	
Total	19,143.5	**538**		

11.48 is well past 3.01, so we reject H_0

At least one center's mean differs at endline. F doesn't say which one

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

The test, and what comes after

Steps 1 and 2: set up the test

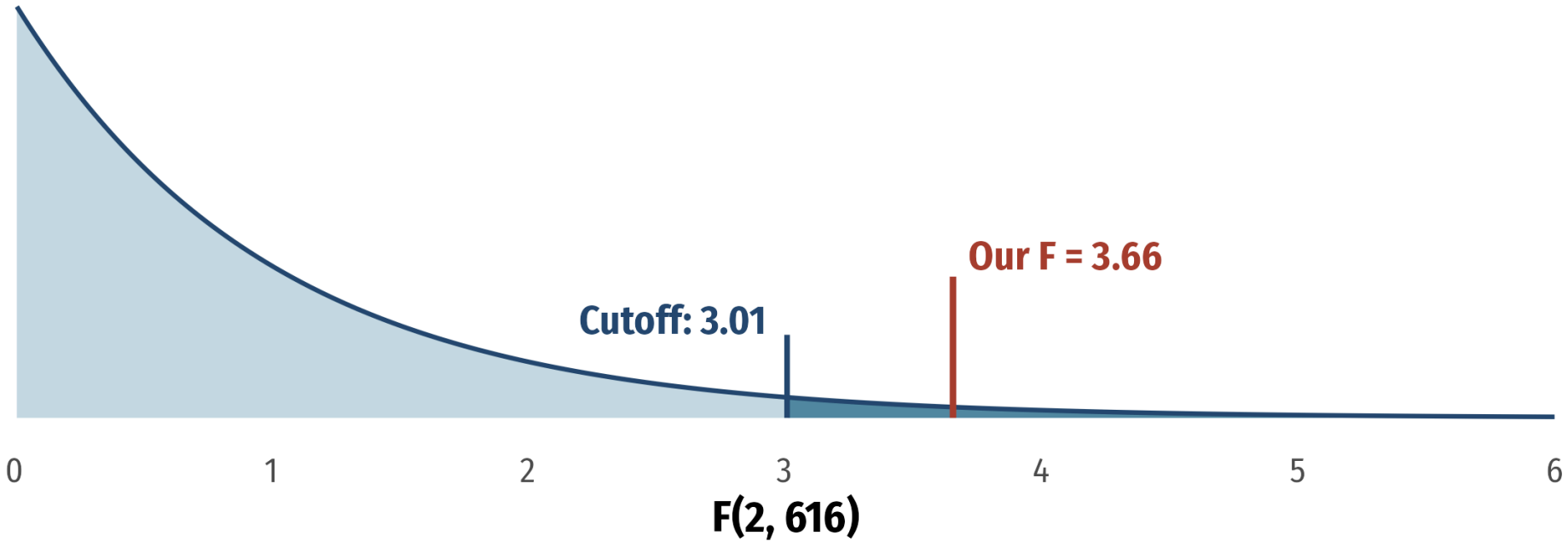
At baseline, do math scores differ across the three centers?

Step	In words	In symbols
1. State H_0	The three population means are the same	$\mu_{Chhatarpur} = \mu_{Govindpuri} = \mu_{Tekhand}$
2. Set the criteria	$\alpha = 0.05$, and two df	df = 2 and 616
	Reject H_0 if F is beyond the cutoff	F(2, 616) beyond 3.01

There's one cutoff. Differences in any direction make F bigger, so H_1 needs no direction

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

Steps 3 and 4: F passes the cutoff



3.66 is beyond 3.01, so we reject H_0. p = 0.026

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

How big? About 1.2% of the total

$$\hat{\eta}^2 = \frac{SS_{\text{between}}}{SS_{\text{total}}} = \frac{110.5}{9419.9} = 0.012$$

At baseline, math scores differed across the three centers, F(2, 616) = 3.66, p = 0.026, $\hat{\eta}^2$ = 0.012

Real, but small. And no lottery made these groups, so it's a difference, not an effect

In the book: η^2 (eta squared), with no hat. Regression calls this share r^2

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

With two groups, F is just t^2

Mindspark and control at endline	Topic 9: t-test	Today: ANOVA
Test statistic	t(537) = 5.371	F(1, 537) = 28.85
Cutoff	±1.964	3.859
p	Below 0.001	Below 0.001

$5.371^2 = 28.85$, and $1.964^2 = 3.859$. Same test, same decision

Squaring folds t's two tails into F's one. ANOVA is the t-test, opened up to more groups

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

F doesn't say which group differs

With three or more groups, a significant F doesn't name the group

One center against the other two?

All three apart?

So after a significant F, we compare pairs. These are called post hoc tests

That's our t-tests again, so the false alarms are back. Post hoc tests set a stricter α

Bonferroni: split α across tests

$$\alpha_{\text{adjusted}} = \frac{\alpha}{\text{tests}} = \frac{0.05}{3} = 0.0167$$

Test each pair at 0.0167, not 0.05

Overall: $1 - (1 - 0.0167)^3 = 0.049$

That's just under 0.05, as we wanted. Bonferroni plays it a little safe

Try it: Chhatarpur against Tekhand

Chhatarpur: n = 161, M = 11.67. Tekhand: n = 376, M = 10.78. MS within = 15.11

MS within is our pooled variance. Find the standard error of the gap, then t

With $\alpha = 0.0167$, the cutoff is $t(616) = \pm 2.40$. Decide: reject or maintain?

Work alone, compare with a neighbor, then we vote

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

Try it: Answer

$\hat{\sigma}_{M_1-M_2} = \sqrt{(15.11 / 161 + 15.11 / 376)} = 0.366$, so $t = 0.89 / 0.366 = \mathbf{2.43}$: reject H_0

Pair	Gap	t	p	At $\alpha = 0.0167$
Chhatarpur and Tekhand	0.89	2.43	0.015	**Reject H_0**
Govindpuri and Tekhand	0.82	1.72	0.086	Maintain H_0
Chhatarpur and Govindpuri	0.07	0.14	0.890	Maintain H_0

Only one pair clears the stricter bar. Tukey's test agrees: p = 0.041

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

The blueprint so far

Blueprint step	Statistic	Bottom of the fraction	df
1. One score	z	σ	None
2. One mean, σ known	z	σ_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**	$\mathbf{MS_{within}}$	**k − 1 and N − k**

Same idea every time: what we saw, over what chance alone would give

What you can do now

Say why many t-tests inflate the risk of a Type 1 error

Explain F as a ratio of two variances, and why it averages about 1 under H_0

Split SS and df, then build a source table

Run the four steps for an F-test, and report it with $\hat{\eta}^2$

Use a Bonferroni correction to find which groups differ

Next steps

Friday, February 26: Lab 8

A 15-minute individual check on two samples, then gains and ANOVA. Bring a calculator and a laptop

Tuesday, March 2: Topic 11

Correlation. Read sections 14-1 to 14-4, and skim 14-5

Sunday, February 28, 6 PM: Chapters 11 and 12 are due in MindTap