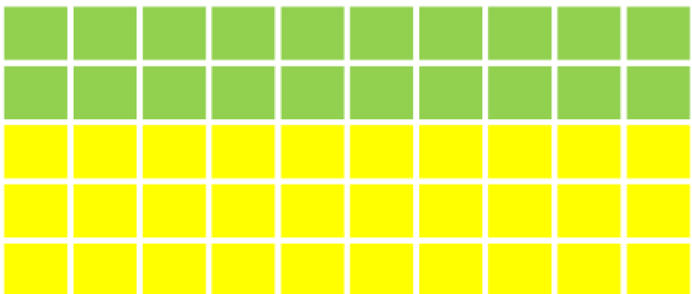
Simpson’s paradox 1

(by Filippo Simini)

“Gerrymandering” consists of the manipulation of the boundaries of constituencies in order to alter the electoral results in a non-proportional system.

For example, consider the following region where each square represents a precinct: green squares vote for party A and yellow squares vote for party B. Out of the total 50 precincts, 20 (40%) vote for party A, and 30 (60%) for party B.

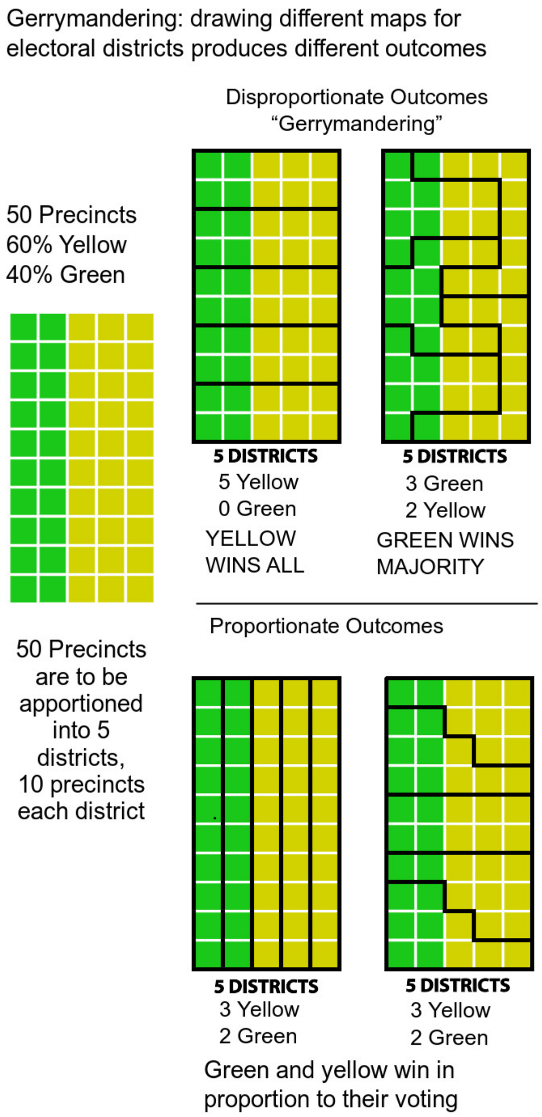


Can you draw 5 constituencies of equal size (10 neighboring precincts each) so that party A and B win in proportion to their overall voting?

Can you draw other 5 constituencies of equal size (10 neighboring precincts each) so that party A, the minority party, wins in the majority of constituencies?

Solution

From https://en.wikipedia.org/wiki/Gerrymandering:



Simpson's paradox 2

(by Filippo Simini)

Gerrymandering is a real life example of what's know as Simpson's paradox, where a trend appearing in different groups disappears when the groups are combined. Another example is the following:

A school has two classes. Class 1 has 25 students, 11 males and 14 females, and class 2 has 23 students, 13 males and 10 females.

Both classes take the same Maths test.
Overall, female students did better than male students: 12 out of 24 females passed the test (success rate 0.5), while just 11 of the 24 males passed it (success rate 0.458).

However, in each class, males had a higher success rate than females!
How is this possible?

Can you find a set of results such that male students have a higher success rate than female students in each class, but a lower success rate overall?

	Class 1	Class 2	Total
Males	**? / 11**	**? / 13**	11 / 24
Females	? / 14	? / 10	**12 / 24**

Solution

A possible solution is

	Class 1	Class 2	Total
Males	**9 / 11**	**2 / 13**	11 / 24
Females	11 / 14	1 / 10	**12 / 24**

Let m_i (f_i) be the number of male (female) students passing the test in class i.
In general, a set of scores for which males have higher success rates in each class must satisfy these conditions:

$$m_1 + m_2 = 11 \quad \Rightarrow \quad m_2 = 11 - m_1 \tag{1}$$
$$f_1 + f_2 = 12 \quad \Rightarrow \quad f_2 = 12 - f_1 \tag{2}$$

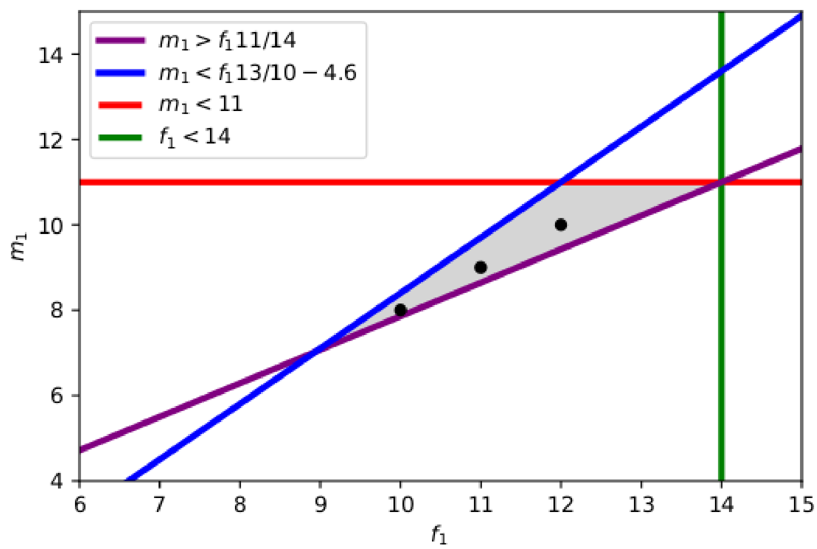
and

$$m_1/11 > f_1/14 \quad \Rightarrow \quad m_1 > f_1 11/14 \tag{3}$$
$$m_2/13 > f_2/10 \quad \Rightarrow \quad m_2 > f_2 13/10 \quad \Rightarrow \quad m_1 < f_1 13/10 - 4.6 \tag{4}$$

plus the constrains on the total number of males and females in each class

$$m_1 < 11 \tag{5}$$
$$f_1 < 14 \tag{6}$$

The solutions of the system lay in the grey region showed below:



Simpson's paradox 3

(by Filippo Simini)

Simpson's paradox has relevant implications on our ability to understand the results of scientific experiments, for example in medical studies.

Consider the previous example, with the same outcomes, but instead of class tests assume that the results describe two independent investigations (test 1 and 2) on the effectiveness of two drugs, A and B. The numerators now correspond to the number of patients successfully treated using the drug.

	Test 1	Test 2	Total
Drug A	**9 / 11**	**2 / 13**	11 / 24
Drug B	11 / 14	1 / 10	**12 / 24**

According to both independent experimentations (tests), drug A is more effective than B. However, when the results of the tests are combined, we reach the opposite conclusion: drug B works better than A.

What should be trusted, the unanimous conclusions of the independent tests, or the reverse indication of the aggregate data?

Solution

Ask the students how they would interpret the results of the experiments and which drug they would give to a patient.

To stimulate the discussion, propose the following variations of the experiment:

1. Suppose that the number of success cases of drug A in Test 2 is 1 instead of 2. Hence, the success rate becomes $1/13 < 1/10$, so drug B becomes more effective in Test 2. Should you prefer drug B over drug A, in this case?

2. Suppose you are told that the participants of Test 1 are all over 40 years old, while the participants of Test 2 are all below 40.
   If you know that a patient is below 40, then using the results of Test 2 obtained testing patients in this age class, you would recommend drug A. Similarly, if you know that the patient is above 40, using the results of Test 1 for old patients, you would also recommend drug A. But if you do not know the patient's age, using the combined data of all tests you would recommend drug B.
   So, if you don't know the patient's age you would recommend drug B, while if you know their age you would always recommend drug A, irrespective of the age!

Simpson's paradox 4

(by Filippo Simini)

Suppose we test the drugs on groups that are ten times bigger than the previous ones. For example, in Test 1 drug A is now tested on 110 individuals instead of 11.
We also assume that the success rates do not depend on the group sizes.

	Test 1	Test 2	Total
Drug A	? / 110	? / 130	? / 240
Drug B	? / 140	? / 100	? / 240

Does increasing the number of participants in all groups resolve the paradox?

Solution

Assuming that the success rates do not depend on the group sizes, the number of successful cases in each group must also become ten times bigger, so that the success rates do not change.

	Test 1	Test 2	Total
Drug A	**90 / 110**	**20 / 130**	110 / 240
Drug B	110 / 140	10 / 100	**120 / 240**

You can see that the success rates of the two experiments combined (in the "Total" column) do not change. Hence, increasing the number tests in each class in the same proportion does not resolve the paradox.

Simpson's paradox 5

(by Filippo Simini)

Suppose we test the drugs on groups of equal sizes, of 100 patients each:

	Test 1	Test 2	Total
Drug A	? / 100	? / 100	? / 200
Drug B	? / 100	? / 100	? / 200

Assume again that the success rates do not depend on the group sizes.

Does considering groups of equal sizes resolve the paradox?

Solution

Let's compute the success rates in each group:

	Test 1	Test 2	Total
Drug A	$\mathbf{9/11 = 0.82}$	$\mathbf{2/3 = 0.15}$	
Drug B	$11/14 = 0.79$	$1/10 = 0.1$	

We see that now drug A is the most successful separately in both tests *and* when all data are combined:

	Test 1	Test 2	Total
Drug A	**82 / 100**	**15 / 100**	**97 / 200**
Drug B	79 / 100	10 / 100	89 / 200

Hence, choosing groups of equal sizes can resolve the paradox.
The first table shows that the overall success rate of drug A can take any value between 0.15 and 0.82, depending on the relative sizes of the groups treated with drug A in Tests 1 and 2. In particular, calling N_i the number of individuals participating in Test i, the overall success rate of drug A will be ~ 0.82 if $N_1 \gg N_2$ (N_1 much larger than N_2) and vice-versa, it will be ~ 0.15 if $N_1 \ll N_2$.
The same applies to drug B.
So, if we expect that groups belonging to Test 1 and 2 should have equal sizes, we should compute the overall success rates as the average rates of all tests.
When groups tested in Tests 1 and 2 do have different sizes (because, for example, there are more young patients than old patients), then we should compute the overall success rates of drugs A and B as weighted averages, where weights are N_1 and N_2 for Tests 1 and 2 respectively.

In general, to avoid ambiguous results, in each Test we should treat the same number of patients with the two drugs. Indeed, for any N_1 and N_2, drug A has a higher overall success rate than drug B:

$$\text{total success A} = N_1 r_{A,1} + N_2 r_{A,2} > N_1 r_{B,1} + N_2 r_{B,2} = \text{total success B} \tag{7}$$

$$= N_1(r_{A,1} - r_{B,1}) > N_2(r_{B,2} - r_{A,2}) \tag{8}$$

where $r_{A,1} = 0.82 > 0.79 = r_{B,1}$ and $r_{A,2} = 0.15 > 0.1 = r_{B,2}$, hence $(r_{A,1} - r_{B,1}) > 0 > (r_{B,2} - r_{A,2})$.