Averages: Still Flawed

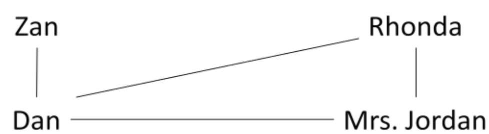
The Friendship Paradox

Supplement for the break after the fourth segment

This supplement is intended to help the discussion around The Friendship Paradox in the ***Averages: Still Flawed*** module.

We will start with the example presented in the video segment. A graph of the original group and the questions asked are included first:

A graph of the group of friends:



Each line means "is friends with"

Questions for the Friendship Paradox:

1) How many friends does each member of this group have, on average?

2) Averaging over each person in this group, what is the average number of friends that each of her or his friends has?

Using tables like the ones presented here, it is hopefully clear how to calculate the two different averages:

	Zan	Mrs. Jordan	Rhonda	Dan
Number of friends in this group	1	2	2	3

1) How many friends does each member of this group of people have, on average?

$$\frac{1+2+2+3}{4} = 2$$

	Zan	Mrs. Jordan	Rhonda	Dan
Number of friends of each friend in this group	3 (Dan)	2 (Rhonda)	2 (Mrs. Jordan)	1 (Zan)
		3 (Dan)	3 (Dan)	2 (Mrs. Jordan)
				2 (Rhonda)

2) Averaging over each person in this group, what is the average number of friends that each of her or his friends has?

$$\frac{3+2+3+2+3+1+2+2}{8} = 2.25$$

Also, as mentioned in the Teacher's segment, we would recommend looking into the extremes of the situation.

Extreme 1: Say Joe is one of the students, and everyone is friends with Joe and no one else. Ask the students to draw the friendship graph over a circle of the names of the people in the group: one straight line from each person to Joe. So, with the exception of Joe, each person has one friend, but he has N-1 friends.

The average number of friends that people have in this scenario is

$$\frac{1*(N-1)+(N-1)*1}{N}=\frac{2*(N-1)}{N}=2*(1-\frac{1}{N})$$

The above equation in words is

$$\frac{1*(N-1)+(N-1)*1}{N}=\frac{Joe's\ friends+Everyone\ else's\ friend, meaning\ Joe}{the\ total\ number\ of\ people}$$

On the other hand, the average number of the friends of friends in this scenario is

$$\frac{1*(N-1)*1+(N-1)*1*(N-1)}{1*(N-1)+(N-1)*1}=\frac{1+(N-1)}{2}=\frac{N}{2}$$

Similarly, this equation in words is

$$\frac{1*(N-1)*1+(N-1)*1*(N-1)}{1*(N-1)+(N-1)*1}=$$

$$\frac{Joe's\ friends\ (N-1)\ each\ have\ 1\ friend+Everyone\ else\ has\ 1\ friend,\ Joe, who\ has\ N-1\ friends}{The\ count\ of\ Joe's\ friends\ \left(1*(N-1)\right)+\ the\ count\ of\ everyone\ else's\ friends\ ((N-1)*1)}$$

In this extreme scenario, the average number of friends converges to a constant, 2. The average number of friends that the friends have, on the other hand, grows linearly with N.

Extreme 2: Assume everyone is a friend with everyone else. Ask the students to draw a fully connected graph. For N students, the average number of friends that people have is

$$\frac{N*(N-1)}{N}=N-1$$

The average number of friends that your friend(s) have in this case is

$$\frac{N*(N-1)*(N-1)}{N*(N-1)}=N-1$$

In this case, the averages are equivalent (e.g. N-1) and there is no paradox.

In-class Activity: Between the Extremes

The activity we recommend for this break looks at scenarios in between the two extremes.

We would recommend groups of approximately 6 students. To make the group of "friends" random and not contingent on social dynamics in the class, you could use the following process to create each group's friendship graph. For each pairing of two people in the group, flip a coin. If it is heads, then they are friends, but if it is tails, then they are not friends. After building a graph like ours, have them fill out a table or do whatever they would prefer to calculate the averages for the two questions.
One important note: it is possible for the coin flips to split the large group into two or more independent groups of friends. That's ok! The process will still work out, and they should use the numbers from the original group and not the subgroups.

Applications of The Friendship Paradox

The following are excerpts from "Friends You Can Count On" by Steven Strogatz, available as of January 2014 at: http://opinionator.blogs.nytimes.com/2012/09/17/friends-you-can-count-on/?hp&_r=0

Application 1: "In a study conducted at Harvard during the H1N1 flu pandemic of 2009, the network scientists Nicholas Christakis and James Fowler monitored the flu status of a large cohort of random undergraduates and (here's the clever part) a subset of friends they named. Remarkably, the friends behaved like sentinels – they got sick about *two weeks earlier* than the random undergraduates, presumably because they were more highly connected with the social network at large, just as one would have expected from the friendship paradox. In other settings, a two-week lead time like this could be very useful to public health officials planning a response to contagion before it strikes the masses.

And that's nothing to sneeze at."

Application 2: "Besides suggesting a strategy for detecting infection, the friendship paradox suggests a strategy for combating it. The idea is to immunize the friends of random nodes, rather than the nodes themselves. See R. Cohen, S. Havlin, and D. ben-Avraham, "Efficient immunization strategies for computer networks and populations," Physical Review Letters, Vol. 91, No. 24 (2003), 247901. Using computer simulations, the authors find that this approach is much more effective than random immunization at halting an epidemic. The technique achieves herd immunity when around 20 to 40 percent of the friend population is immunized, as opposed to the 80 or 90 percent coverage needed when the population at large is immunized."

For those interested in reading the paper referenced in the second excerpt, it is available here (as of January 2014): http://polymer.bu.edu/~hes/networks/chb03.pdf

(Advanced) The Math Behind The Friendship Paradox

The math required to see why the Friendship Paradox works is fairly advanced, as it involves random variables and taking expectations.

We'll start with the definition of variance, which for a random variable X is

$\mathrm{Var}(X) = E[(X-E(X))^2]$

Expanding this, we find

$\mathrm{Var}(X) = E[X^2 – 2XE[X] + [E(X)]^2]$

$\mathrm{Var}(X) = E[X^2] – 2E[X]E[X] + [E(X)]^2$

$\mathrm{Var}(X) = E[X^2] – [E(X)]^2$

Rearranging, we find that

$$\frac{\mathrm{E}[\mathrm{X}^2]}{\mathrm{E}[\mathrm{X}]} = \mathrm{E}[\mathrm{X}] + \frac{\mathrm{Var}(\mathrm{X})}{\mathrm{E}[\mathrm{X}]}$$

E[X] is simply the average number of friends in a group of people represented by X.

What may not be immediately clear is that $\frac{\mathrm{E}[\mathrm{X}^2]}{\mathrm{E}[\mathrm{X}]}$ is the other average, i.e. the average number of the friends of friends. If you look back at the calculation of our original group, you will see the pattern:

$$\frac{3+2+3+2+3+1+2+2}{8} = \frac{3^2+2^2+2^2+1^2}{3+2+2+1} = \frac{(3^2+2^2+2^2+1^2)/4}{(3+2+2+1)/4} = \frac{\mathrm{E}[\mathrm{X}^2]}{\mathrm{E}[\mathrm{X}]} = 2.25$$

So, the average number of friends of friends $\frac{\mathrm{E}[\mathrm{X}^2]}{\mathrm{E}[\mathrm{X}]}$ equals the average number of friends $\mathrm{E}[\mathrm{X}]$ plus the additional term $\frac{\mathrm{Var}(\mathrm{X})}{\mathrm{E}[\mathrm{X}]}$. By definition, the variance of X is always greater than or equal to 0, and thus, except for the "Extreme 2" case presented earlier when the variance is 0, the average number of friends of friends will indeed be larger than the average number of friends in a group of people.