AGAINST ALL ODDS
EPISODE 31 – "ONE WAY ANOVA"
TRANSCRIPT

FUNDER CREDITS

INTRO

Pardis Sabeti
Hello, I'm Pardis Sabeti and this is *Against All Odds*, where we make statistics count.

Today we have a challenge. See if you can guess how much money is in this vase. It's a pretty random mix of quarters, dimes, nickels and pennies – and to give you a hint, I'm pouring the last ten dollars in now...

Jot down your guess, and at the end of the segment we'll let you know how well you did. No prizes I'm afraid.

> **John Kelley**
> I think if we get to 250 or so, we should be fine…

But there *is* a prize, the vase and its contents, for whichever student taking our challenge at Endicott College in Massachusetts gets closest to the right answer.

> **John Kelley**
> What we are trying to demonstrate is something called the "Wisdom of the Crowd," which is that the average of all of the guesses will probably be more accurate than most of the individual guesses. The reason is that everybody has some experience about how much money that might be, and some are gunna guess too high, some are gunna guess too low, and it will kind of average out, we hope, to show that the average will be better than most of the individual guesses.

Pardis Sabeti
Now, John Kelley, a professor at Endicott, and his student Sarah Lovell, aren't being totally up front about the purpose of the experiment. A little harmless deceit is common in psychology, where knowing the real aim of an experiment would negate it. Sure, John and Sarah are interested to see if the wisdom of the crowd holds up, but to get at the real goal of the study we need to take a closer look at the clipboards the students are holding.

The clipboards are the type with storage compartments for extra paper. One of them will have just a sheet or two in back for the duration of the experiment. It weighs just over a pound.

> **John Kelley**
> Got about…just about 1.1 pounds.

Pardis Sabeti

A second clipboard will always be half full of paper, and weigh just over two pounds, while the third is full of paper and weighs just over three pounds.

Each student only ever gets to hold one of the clipboards so they are unaware of the weight difference. And so that the students aren't tempted to rest their clipboards – especially the heavier ones – on the table…

> **John Kelley**
> The only thing we ask you to do is not to touch the table because it is unstable and it could tip over. Okay?
>
> The theory is that the differently weighted clipboards will unconsciously bias them to give different estimates, so that people with the heavier clipboard will give a higher estimate and those with the lighter clipboard will give a lower estimate.
>
> **Student 1**
> $75 and 50 cents.
>
> **Student 2**
> Oh gosh, I was off. I guessed $380. I don't know…
>
> **John Kelley**
> Well we don't…nobody knows how much is in there. One of them might be right. Who knows?

Pardis Sabeti
John Kelley based his experiment on a published study, in which people holding differently weighted clipboards did estimate quantities differently. Studies like this have become popular in experimental psychology. The idea they are testing is that physical experience can influence our thinking in ways we are unaware of.

So for instance, one study showed holding a cup of warm coffee, as opposed to a cold one, made people perceive others as warmer, more generous, more caring. And it gets weirder. Another study showed putting together a simple puzzle made of pieces covered in sandpaper, as opposed to ones with smooth pieces, caused subjects to make harsher judgments about others. Yet another study had people handling a hard object, like a Rubik's cube, or a soft one, like a piece of blanket. The hard object made subjects more likely to judge others as rigid and inflexible than did handling the blanket.

So let's check in with John and Sarah's study. Over the course of a couple of mornings they gave well over 200 students a chance to win the money. A third holding the lightest clipboard, a third the medium weight one, and a third the heaviest. Estimates varied wildly.

Student Guesses
45 bucks, 62 cents.

$75 and 63 cents

$95 and 7 cents.

I said $112.

I guessed $63 and 75 cents.

I guessed $72 and 50 cents.

$19 and 55 cents.

$99

John Kelley
$99?

Student Guesses
$777 and 77 cents.

$178

John Kelley
And 20 cents.

Student Guesses
…and 20 cents.

I guessed $529 and 61 cents.

I guessed $86 and 77 cents.

John Kelley
The range of choices, or guesses that people made is enormous, from very little, as little as $10, to as much as $1,500. So we have gigantic amounts of variance and if our effect is relatively small, you know, as compared to the error variance it's going to be very hard to detect.

Pardis Sabeti
John and Sarah enter the guesses, grouped according to the weights of the three clipboards, then computed the average guess in each of the three groups.

John Kelley

If you just look at the means, it's very promising. We've got, for the light clipboard, $107. With the medium-level clipboard, it's about $130. If you look at the heavy clipboard it's $143, so we have a very nice linear effect, exactly what we predicted. If you plot that with a graph, that I can do right like this, you see this very nice beautiful linear effect, and you think, hallelujah, we got it.

However, if you then do a boxplot, you look at the actual distributions, you see there are extreme outliers. And by chance, probably, in Groups 2 and 3, there's more outliers than in Group 1, so it's possible that this linear effect is driven by just a few people, it might just be a fluke.

Pardis Sabeti

The most striking thing about John's data is the extreme variance within each group as compared to the variance between the means of each of the groups. In other words, the noise is threatening to swamp any signal that might be there. To see if there is a meaningful signal – if in fact the variance between the population mean of the three groups is significantly different – he turned to a technique called a one-way Analysis of Variance, or ANOVA. In asking the question – is there a significant difference between the population means of three or more groups? – a one-way ANOVA actually sets up the hypothesis that there is in fact *no* significant difference – the so-called null hypothesis, where the population means of the three groups are really the same.

John Kelley

It's kind of a weird thing that we test the null. You wonder, why don't we test the actual hypothesis. The reason is that the hypothesis is non-specific. We have a vague hypothesis that heavier clipboards should result in higher estimations, but we don't have any idea really about how big that effect is, so it's very hard to generate probabilities when we have this kind of squishy kind of hypothesis. The null hypothesis is that all the means in the population are the same, so in other words, if we did this infinitely the averages in the three groups would really be the same, there's absolutely no effect of the weight on people's estimations.

Pardis Sabeti

Of course what John is hoping for is sufficient evidence to reject the null hypothesis. To do that he runs an ANOVA using software to compute a statistic called *F*.

The numerator measures how much variability there is between the sample means, while the denominator measures how much variability there is within the

individual observations. In John's case, F equals 0.796. To reject the null hypothesis, the probability of obtaining a value at least as extreme as $F = 0.796$ needs to be less than 5% – in other words a p-value less than .05, the usual statistical standard. But the actual p-value John finds is .45.

> **John Kelley**
> So in this case, since our p-value is .45, there's a 45% chance of getting this big a difference, or bigger, just by chance. So at this point, we cannot say that we have by any means shown that this hypothesis is true.

Pardis Sabeti
One of the underlying assumptions of ANOVA is that the data in each group are normally distributed, which is not the case here given the skewed nature of the data. John's students tried some statistical manipulations on the data to make it more Normal, and then re-ran the ANOVA. However, the conclusion remained the same – no evidence that the means of the three groups are different.

But what if the data had looked like this instead? The sample means are the same: $107, $130, and $143, but this time the data are less spread out about those means. In this case, the value of F turns out to be 33.316 and the corresponding p-value is essentially zero. Our conclusion in this hypothetical case would be to reject the null hypothesis—the population means *do* differ. And at that point we could turn to different statistical tools to explore the differences further.

But, sadly, in John Kelley's study, the harsh reality of a rigorous statistical analysis has shot down the idea that holding something heavy causes you, unconsciously, to make larger estimates – at least in this particular study.

> **John Kelley**
> Inevitably you have to be disappointed because you put a lot of work into it and you find it didn't work. And there's actually reasonably good evidence that a lot of studies are wrong, that it's going to happen. A reason for that would be – imagine if this was me and I was continuing to do this kind of work and I was the first one to do it – I'm not, but if I was the first one to do this particular kind of study, I might keep working at it and try this and try that and try another thing and eventually I might get it to work. Now that might mean that it does work under those conditions or it might mean that I just got lucky eventually.

Pardis Sabeti
But if the real experiment didn't work, what about the cover story that John and Sarah were testing – the idea of the Wisdom of the Crowd? The actual amount in the vase is $237.52. Here's a histogram of all the guesses. You can see instantly that most people were way too low, almost half of them being under $100. The mean of the estimates is $129.22 – more than $100 off, but still better than about

three-quarters of the individual guesses. So the crowd was wiser than the people in it – but still not all that wise.

Unlike the actual winner, Chris Johnson, who turned out to be an accounting major who's worked in a bank.

>**John Kelley**
>Congratulations

>**Chris Johnson**
>Thank you.

Pardis Sabeti
His estimate was $235.

>**John Kelley**
>You were within $2.52. So we are going to give you this coin jar. So literally. So here you go. It's really heavy.

>**Chris Johnson**
>Thank you.

>**John Kelley**
>That's for you to keep and I guess you go to your bank and do some more counting, essentially.

Pardis Sabeti
So how well did you do? We borrowed back the vase for this taping, and to give you one last chance to compare your guess with the reality of $237.52. Now the jar and coins are on their way back to our winner.

I'm Pardis Sabeti for *Against All Odds*. See you next time!

PRODUCTION CREDITS

Host – Dr. Pardis Sabeti

Writer/Producer/Director – Graham Chedd

Producer – Maggie Villiger

Associate Producer – Katharine Duffy

Editors – Dave Berenson
–Seth Bender

Director of Photography – Dan Lyons

Sound Mix – Richard Bock

Animation – Jason Tierney

Title Animation – Jeremy Angier

Web + Interactive Developer – Matt Denault / Azility, Inc.

Website Designer – Dana Busch

Production Assistant – Kristopher Cain

Teleprompter – Kelly Cronin

Hair/Makeup - Amber Voner

Music
DeWolfe Music Library

Based on the original Annenberg/CPB series *Against All Odds*,
Executive Producer Joe Blatt

Annenberg Learner Program Officer – Michele McLeod

Project Manager – Dr. Sol Garfunkel

Chief Content Advisor – Dr. Marsha Davis

Executive Producer – Graham Chedd

FUNDER CREDITS