

Does it really matter – bad tribes & bad vibes

If you've ever experienced the joy that is watching the reality competition TV show *Survivor*, you'll understand that there are certain constructed ideas among contestants and viewers about how to play the game. If you're into maths (like many of the fanbase and presumably readers of this essay) you'll understand the draw to statistical analysis of these ideas. I hope through my writing I'll be able to guide you through one of the most core *Survivor* principles and debunk the 'make or break' myth, while also teaching you a little bit of basic data science.

What is Survivor?

Over the years since it first aired, *Survivor* has gone through many evolutions from the classical survival show to become the highly social and strategic game it is today. Nowadays there are several spin offs from the original American version including Australian, British and South African and New Zealand versions. This allows for an immense pool of data to draw from and a hope of counteracting the twists and turns that are dotted through reality TV.

At the start of each season, contestants are split into Tribes (usually two, sometimes three). Tribes then compete for rewards (food and comfort items) as well as immunity. Failing to win immunity sends the other tribe to vote one of their own tribemates off the show (tribal council). Around halfway through the game, tribes merge into one tribe and contestants then compete for individual reward and immunity (winning immunity protects you from elimination). After merge, contestants no longer leave the game entirely, instead they sit on a Jury who eventually vote for the winner between the final two or three.

What is 'the make or break' principle?

Before you rush to google, you should know this is something I've named in order to try and label the shared idea between contestants that the first episodes tend to set the stage for the rest of the season and so are *perceived* as being very crucial.

The first few episodes introduce the contestants, and they have their first go at challenges and tribal councils. The first alliances are formed, and the weakest are singled out. Naturally, after months of selection to be on the show, contestants are under a lot of pressure to make it in their tribes – unfortunately not all succeed.

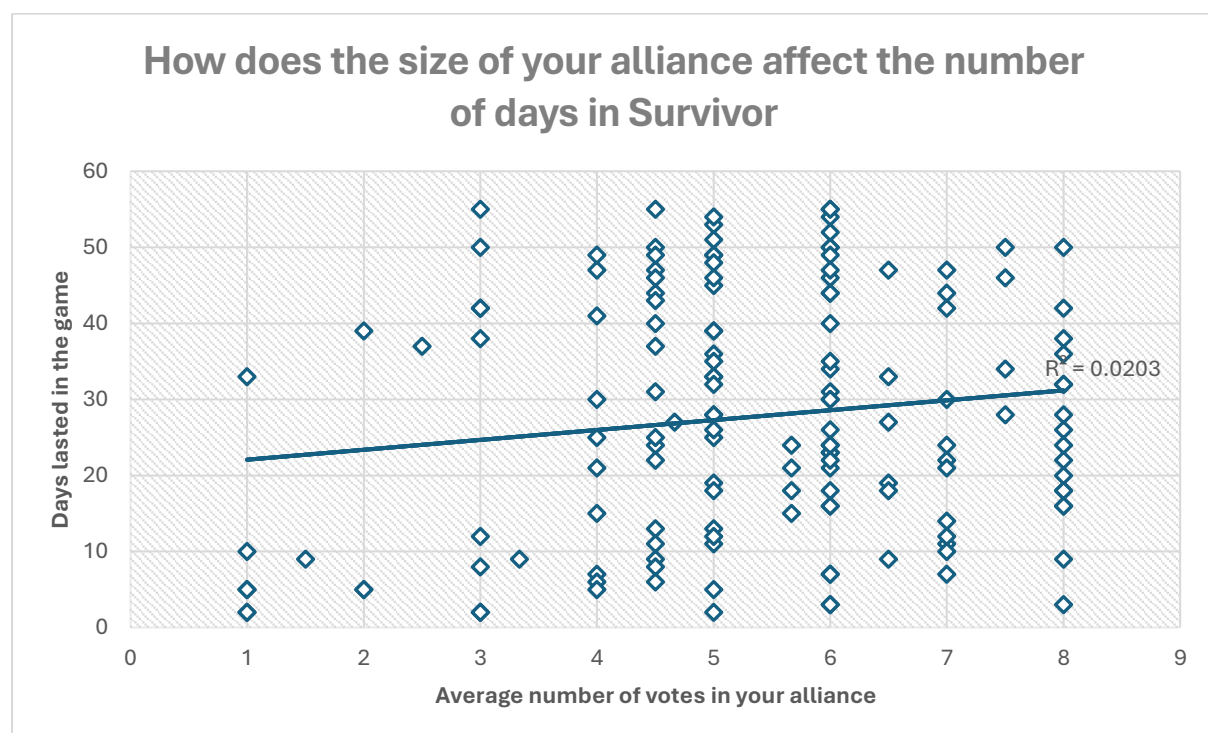
My curiosity was in looking at if this really mattered. Does starting on the back foot actually affect your long-term trajectory?

For this essay, I've decided to look at two aspects: initial alliances and initial challenge success. Alternatively, I could have looked at screen time, number of confessionals, however these can be more easily skewed by producers.

Initial Alliance sizes

The vast majority of data from tribal councils that you will find online centres on what's called a 'vote for bounty'. This refers to whether or not the contestant voted for the person who actually ended up going home. I had some problems with this – firstly it completely ignores the certainty of the vote (as in how easily it could have gone another way: was it straight forward or was there a close split?). For example, in a vote where five contestants vote one way, and four the other, the four who were on the wrong side of the vote would get a VFB (vote for bounty) of zero. That's saying they were 100% wrong, which is not accurate. Secondly, it assumes if you didn't vote for someone you didn't want them gone. This will sound very confusing to anyone who has not watched the show but it's much simpler than it sounds. Basically, you may decide in your alliance to split a vote so that you have a back up in case your first target plays an immunity idol (causing them to be protected). The problem here is that everyone who voted for the backup would get a VFB score of zero – even when they were playing well?

In order to try and correct the blind spots in this method, I focused on a different aspect: how many people voted the same way as you. I started by looking at the Australian series for seasons 1 – 8. This is because I have access to a large sample size, without the worry of constant twists that are prevalent in the American series. For each contestant I recorded their days lasted and plotted it against average alliance size.



On the x axis is the average number of votes cast during the three tribal councils that were the same as the player. On the y axis is the days lasted by the player. For example let's look at Flick for season 8: Her tribe won the first immunity challenge so didn't vote, in the second tribal council she voted for Janelle alongside 6 others, and in the third tribal council she voted for Gavin alongside 5 others. Flick lasted 47 days in the end.

Mean number of votes in alliance = $(7 + 6)/2 = 6.5$

Therefore point plotted (6.5, 47)

What can we learn from this data? If you've got good eyes, you might have noticed that I've noted the R^2 value. This is the coefficient of determination and it is a measure of the proportion of variance that is explained by the independent variable. In other terms, this is a measure of how much of the data can be explained using the regression model (the line of best fit) and how well the model can predict the outcome.

The coefficient of determination is expressed at a number between 0 and 1 (or 0% to 100%) where 0 indicates no explanatory power (ability to predict) and 1 indicates perfect prediction of the dependent variable.

How is it calculated? There are two ways – either we calculate R and then square or go straight for R^2 .

Let's do the first method. A R value is known as the correlation coefficient and it shows the strength and direction of a correlation. To calculate R:

- 1) Collect paired data for the two variables (Mean Alliance Size, Days lasted)
- 2) Calculate means of the x and y

Meanx = Sum of x/ number of

$$\bar{x} = (\sum x_i) / n$$

$$\bar{x} = 799.6667 / 150$$

$$\bar{x} = 5.331$$

$$\bar{y} = 27.707$$

- 3) Calculate deviations from the mean for each data point

Deviationx = $x - \bar{x}$, Deviationy = $y - \bar{y}$

- 4) Calculate standard deviations for each variable
- 5) Calculate R using the formular

R = Sum of the products of the deviations / product of n by the standard deviations

$$R = \Sigma(\text{Deviationx} * \text{Deviationy}) / (n * \text{SDX} * \text{SDY})$$

$$R = 0.14$$

$$R^2 = 0.02$$

So, what does this mean? An R value of 0.14 tells us that only 14% of data points are due to the independent variable. So for a contestant, the size of your initial alliance will only predict your future outcome about 14% of the time! This is great news if you were planning on competing on the next season because it suggests that you really do have a lot of wiggle room if you do end up the back foot for the first few votes.

Challenges and if you win them

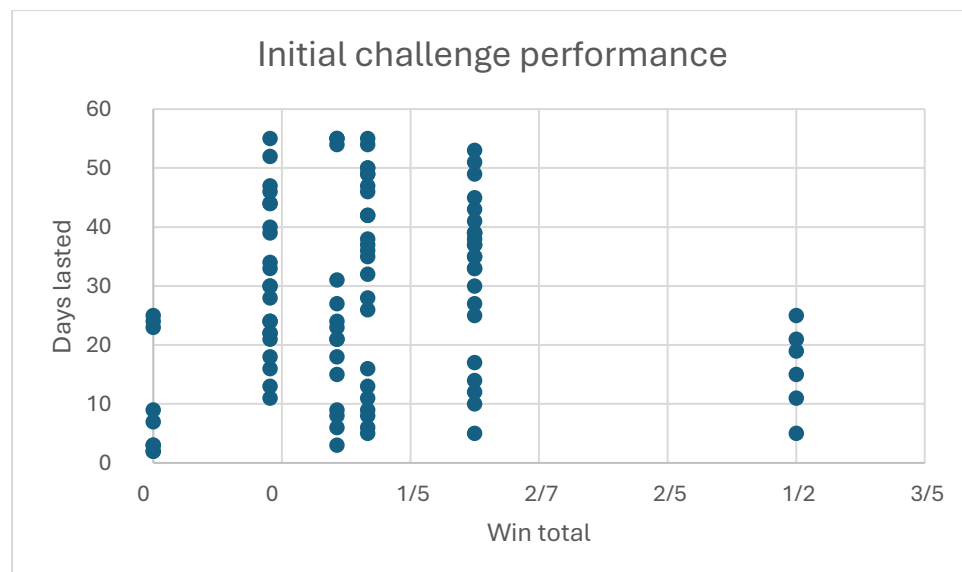
To start off with I would like to say that in my opinion you should of course try and win your first few challenges. There is occasionally the opinion held by some contenders that they should be less competitive in order to not highlight themselves as a threat or that it would be beneficial to lose so that they could cut some dead weight.

The reason I think initial challenge win record would be a useful metric to look at is because theoretically there could be a correlation: stronger tribes win in the beginning, they continue to win, less of their tribe goes out, come merge they have a majority, everyone in their tribe gets further than they originally would.

Inspired by my first success in debunking I thought id give this a go too.

Looking at seasons 1 - 5 of the Australian series I awarded a fraction of the win to each of the members of a tribe. For example, looking at Flick again, her tribe of 8 won the first two challenges but lost the third so she gets a score of $1/8 + 1/8 + 0 = 1/4$.

My results are as follows:



As you can see there is essentially no correlation between the initial challenge wins and how well contestants are to perform later in the season.

Conclusion

I hope this provided some interest, probably something a little different, possibly made you think of something new. Should you be next in line for Survivor, you can sleep well knowing that the first impressions really are not everything despite what other contestants may believe and that there is plenty of room for recovery.