ASA Regression Thinking

In its simplest form, a regression is just a statistical tool that helps you answer this question: "When one thing changes, how much does another thing usually change?" In other words, are the two things related at all and, if they are, what is the strength of that relationship?

In practical terms, imagine this: You notice that when it rains more, people carry more umbrellas. If you then observed numerous days, and made a dot on a graph each day for 1. How much it rained and 2. How many umbrellas you saw, after time, you could draw a line through those dots in a way that linked up the biggest clusters.

That line is a regression.

It would probably tell you: If the amount of rain increases marginally, the number of umbrellas usually goes up a little too; and, if the amount of rain increases significantly, umbrellas usually goes up a lot.

Regressions are also exploratory, you need to test a few things to see if they are related. In our case, we have four key variables we want to test: The direction of ANC, EFF and Other Parties support, and turnout. All with a view to determining their effect on ASA support.

I am not going to set out all the maths here, it is too much, but I will give you a summary of the results.

I was exploring the following question: Where did ASA's 2021 vote in Region D come from?

To do this, I tested four possible influences on ASA's 2021 vote:

- ANC strength in 2016
- EFF strength in 2016
- Other parties' strength in 2016
- Turnout change (2016 to 2021)

I then ran a series of regressions to see what came up.

Just a note here: the strength or credibility of any regression is measured by something called R^2 (r-squared). It tells you how much of the pattern the regression explains, on a scale from 0.00 – the regression explains nothing – to 1.00 – the regression explains everything.

With real-world political data, you are never going to get 100%. Human voting behaviour is too complex. But as a rough guide: at 0.3 you should start paying attention, at 0.5 you have a meaningful association worth taking seriously, and at 0.7 and above you are dealing with something very strong.

The same logic applies to a "coefficient" – the individual variable's score in each regression. The coefficient tells you how strongly that variable is associated with ASA's performance, after controlling for the others. In our case, if strong, it will not prove voters switched because of that variable, but it will tell us the environment in which ASA is most likely to thrive.

All of these regressions, except for the final few restricted ones, were run across all 38 wards.

The first regressions I ran were what we can call "Change Models" – and explored the question: Where party support fell between 2016 and 2021, did ASA do well? I ran regressions using ANC percentage point change, EFF percentage point change, other party percentage point change, and turnout change.

The results were modest. Other party decline showed some association. EFF decline showed a moderate association. ANC decline was weakest. Turnout did play a role. R^2 was not great. Nothing decisive.

At this stage, no single party collapse explained ASA. But the EFF's slightly stronger relationship was noticeable, and worth pursuing.

So next I tried what we can call "2016 Share Models". Here the question changed. Instead of looking at collapse, I asked: was ASA stronger in wards where a party was strong in 2016?

EFF 2016 to ASA 2021 produced a coefficient of about 0.36, with an R^2 of 0.12. ANC 2016 to ASA 2021 showed a very weak relationship.

These were still generally weak results. But again, the EFF stood out. The R^2 was low, meaning much was unexplained, but the EFF coefficient was clearly stronger than ANC or Other Parties. That suggested that ASA's geography might resemble prior EFF geography more than anything else.

Next I moved to an "Absolute Vote Model" and shifted from shares to raw votes. The question here was: in wards where a party had a larger vote pool in 2016, does ASA have more votes in 2021?

Most associations were weak, but again the EFF coefficient – about 0.42 – was significant. R^2 was around 0.30, stronger than before.

In plain terms, this model was saying that, where the EFF had 1,000 more votes in 2016, ASA tended to have about 420 more votes in 2021. This does not mean 42% of EFF voters switched. It means ASA's ward-level strength tracks very closely with where the EFF was strongest.

By now, the pattern was becoming hard to ignore. So I combined the variables in what I consider the most important model – the "Full Combined Model".

Here ASA 2021 votes were regressed on ANC 2016 votes, EFF 2016 votes, Other 2016 votes and turnout change together.

The result: $R^2 = 0.545$. That means this model explains more than half of the variation in ASA's ward-level performance. In electoral analysis, that is strong.

The coefficients were revealing:

- EFF = roughly 0.44–0.47 and statistically significant.
- ANC = roughly 0.14–0.17 and weaker.
- Other parties = near zero.
- Turnout = significant.

When all plausible explanations were allowed to compete in the same model, the EFF remained the largest and most consistent predictor. At this stage, it was increasingly clear that EFF strength in 2016 was the most powerful structural factor associated with ASA's 2021 vote.

I then tried to "force-test" this theory by removing ASA's intercept. Normally, regressions include an intercept – a baseline ASA vote that exists even if all other variables are zero. That intercept is not guessed; it is mathematically derived so that the model's predictions line up with the real averages.

But I removed it and forced the model to assume that every ASA vote must have come from ANC, EFF, Other Parties or as a result of turnout changes.

The result was an R^2 of about 0.96. That is artificially high because we forced the attribution. But what mattered was the ranking of the coefficients:

- EFF = roughly 0.45–0.55.
- ANC = roughly 0.15–0.23.
- Other parties = very small.

Even when forced, the EFF remained the largest contributor. The ordering did not change.

To ensure turnout was not driving the result, I then removed turnout and re-ran the model. The EFF coefficient remained strong, ANC weakened further, and Other Parties remained negligible. The EFF association survived turnout adjustment, which is important.

Finally, I restricted the pool of wards. Up to this point I had used all 38 wards, but there is a risk that uniformity dampens variation (and Soweto's results, across the board, are very uniform, it basically acts like one big ward in general terms). So I told the model to look only at the most extreme wards – the top 10, 8, 7, 6 and 5 wards per support group in 2016 – and re-ran the combined model each time, with turnout included.

The results were remarkably stable:

- Top 10 per party: $R^2 = 0.54$, EFF ≈ 0.47
- Top 8 per party: $R^2 = 0.56$, EFF ≈ 0.48
- Top 7 per party: $R^2 = 0.69$, EFF ≈ 0.54
- Top 6 per party: $R^2 = 0.68$, EFF ≈ 0.55
- Top 5 per party: $R^2 = 0.87$, EFF ≈ 0.54

As the sample narrowed to the most politically intense wards, explanatory power increased, which is expected. But the critical point is that the ranking never changed. In every version, the EFF coefficient remained the largest. So across change models, share models, vote models, combined models, forced models, turnout-adjusted models, and restricted stronghold tests, one pattern persisted:

EFF 2016 strength is the strongest and most consistent structural predictor of ASA's 2021 vote in Region D.

Other, less significant patterns included:

- ANC contributes, but less consistently.
- Other parties explain very little.
- Turnout matters, but it does not overturn the EFF relationship.

This does not prove individual EFF voters switched to ASA. These kinds of regressions cannot tell you that. But at the ward level, geographically and statistically, ASA's vote aligns most strongly with where the EFF was strongest in 2016. And so one can fairly deduce ASA voters are most likely and primarily drawn from the same pool the EFF draws on.

And crucially, that result survives changes in model specification, variable definition, inclusion or exclusion of turnout, inclusion or exclusion of an intercept, and restriction of the sample.

When a relationship survives that many tests, it stops being anecdotal and starts being statistically significant and so it is worth taking seriously.

Up to this point, most of the regressions were asking what explains ASA's rise. But that only tells half the story. If we assume that ASA made major inroads into the ANC vote, then the ANC's own numbers should show that clearly. So I flipped the question and ran regressions on the ANC alone.

The first of these asked a very simple question: where the ANC lost the most between 2016 and 2021, was ASA strongest? In other words, is ANC decline geographically aligned with ASA strength? The answer was largely no. When ANC change was regressed directly on ASA strength, the explanatory power was extremely low and the relationship was not statistically significant. That means wards where ASA was strong were not automatically the wards where the ANC fell the most. If ASA had devastated the ANC in a direct, one-to-one way, we would expect that pattern to be clear. It was not.

I then ran what we can call an ANC "retention" model. Instead of looking at raw decline, I measured how much of its 2016 vote share the ANC managed to retain in 2021, ward by ward, and then tested what explained that retention. This model produced an R^2 of about 0.43, which is meaningful. What it showed was that stronger ASA wards were associated with lower ANC retention. That is important. It suggests ASA presence does correlate with weaker ANC performance. But it also showed that historically EFF-strong wards were associated with higher ANC retention.

Finally, I ran a direct ANC competition model, where ANC 2021 share was explained by ANC 2016 share, EFF 2016 share and turnout change. This model had an R^2 of about 0.58, which is strong. As expected, ANC 2016 strength strongly predicted ANC 2021 strength. Interestingly, EFF 2016 strength was also positively associated with ANC 2021 strength. That means wards that were historically intense ANC–EFF terrain remained strong ANC wards in 2021. Again, that does not look like a clean collapse story. Turnout decline, meanwhile, significantly hurt ANC performance, reinforcing the idea that demobilisation played a role in reshaping outcomes.

So when these ANC-focused regressions are added to the earlier ASA models, the picture becomes more complete. ASA strength is associated with weaker ANC retention, but ANC decline does not line up neatly, ward by ward,

with ASA strength alone. Turnout and prior political structure matter a great deal. At the same time, the earlier finding still stands: ASA's strongest structural alignment remains with prior EFF terrain.

Taken together, this means the story is not simply that ASA ate directly into the ANC in its heartland. Nor is it that ASA emerged out of "Other Parties" space. The evidence points to a more structural reshaping of opposition terrain, in which ASA's rise is most closely associated with prior EFF-type wards, while ANC decline appears to be influenced by both turnout and fragmentation rather than a straightforward one-to-one transfer with ASA.