SupeR Bowl Babies

**Objective**

For the past two years, commercials have suggested that the cities of the Super Bowl-winning team see a rise in births in the nine months or so following a Super Bowl win. This analysis seeks to statistically test that assertion and estimate the magnitude of the effect, if any.

**Analysis**

The key input to this analysis is a CDC-published, county-and-month-level dataset of birth rates. The most notable step of the curation was transforming county-level data into a dataset aggregated to a geographic region appropriate for capturing the effect of a team winning the Super Bowl. This was almost entirely a judgement call. The counties that were rolled into being eligible to pick up the effect of winning can be found explicitly in the code. The principle implemented here, though, was based on the likelihood that people consider themselves from the same place as the locale the team represents.

The second kind of aggregation was based on the dimension of time. I aggregated the monthly data into quarterly data to smooth out some of the noise and analytically recognize that there may be a lagged effect to winning. After all, I am not strictly interested in an effect nine months after the Super Bowl, rather an effect *around* nine months after the Super Bowl.

Once the data were curated, the structure of the initial analysis was to compare two quantities: 1) the actual change in birth rates in the final months of the year, and 2) the expected change in birth rates for that period. In cases where the actual change is greater than the expected change, we have validated that there is in fact an increase in births after a Super Bowl.

It is worth noting that calculating changes in birth rates is a delicate matter. There is a temptation to evaluate the changes in consecutive months, but this approach is misguided given the consistent cycles throughout the year that we observe in the data. Changes, in this analysis, are evaluated relative to the same period in the previous year. This allows the model to ignore the periodicity in the data and focus on level changes in the cycle. The null hypothesis then is that changes in January births are similar to those for April. And that changes in April births are similar to those for November. By extension, the expectation is that the average change in birth rates outside of the post-Super Bowl period is similar to the average change within the period. The graphs available [here](https://github.com/diagdavenport/SuperBowlBabies/blob/master/Results/Year%20over%20year%20differences.pdf) indicate the differences between those averages and how they coincide with Super Bowl participation.

This initial analysis reveals that there are indeed changes in birth rates that coincide with Super Bowl participation and victory, but the directions and magnitudes are not readily apparent by visual inspection. The next phase of the analysis was to take a more refined approach and attempt to rule out spurious correlations and potentially confounding variables.

The chosen solution was a linear regression, modeling changes in the Q4 births rates as a function of:

1. Q1 US GDP (fixed to 2009 dollars),
2. Super Bowl participation (binary flag), and
3. Super Bowl Victory (binary flag)

**Results**

The results were emphatic and surprising, but perhaps more interesting as a result. The statistically significant coefficients were counterintuitive in both direction and magnitude. It turns out that there are two significant effects that need to be quantified separately when considering the effect of Super Bowls on births. The first is that the impact of participating in a Super Bowl generally causes a drag on birth rates later in that year; when teams participate in the Super Bowl, their cities should expect a decrease in birth rates relative to the same time last year. That is, unless they win. The second effect that gets parsed from the model is that, if the team wins, the increase in birth rates offsets the loss from participating—the net effect of winning the Super Bowl is effectively zero (as far as birth rates are concerned)!

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t Value | Pr(>|t|) | Significance |
| participatedTRUE | -0.03 | 0.01 | -2.24 | 0.03 | \*\* |
| wonTRUE | 0.03 | 0.02 | 1.70 | 0.09 | \* |
| poly(GDP..2009.dollars., 3)1 | 0.06 | 0.04 | 1.51 | 0.14 |  |
| poly(GDP..2009.dollars., 3)2 | -0.02 | 0.04 | -0.68 | 0.50 |  |
| poly(GDP..2009.dollars., 3)3 | -0.03 | 0.04 | -0.79 | 0.43 |  |
| Signif. codes:  0.01 ‘\*\*’ 0.05 ‘\*’ |  |  |  |  |  |

**Conclusions**

It turns out that when teams play for the Super Bowl, there is much more at stake than the trophy, but not in the way we might’ve expected. Teams are effectively vying to prevent declines in their city’s birth rate. This has useful public policy implications as hospitals can begin to develop better expectations for pediatric demand after Super Bowls.

**Follow-on Work**

Currently, the biggest vulnerability of the model is that there is no accounting for changes in the population. More precisely, there is no accounting for changes in the population of child-bearing women. Further refinement of the model would benefit from controlling for these sorts of shifts in the population.

Additionally, the model could be improved by refining the economic activity measure (currently US GDP) to be some region-specific measure of economic activity. This lack of regional focus is a possible culprit for the lack of significance for the activity measure.

Finally, in further development, I would also like to test the sensitivity of the model to aggregating to the quarter level.