15-May-2023

USPP-2022-0019 - Fallacies in Statistically-Based Claims about Massive Election Fraud in 2020: A Compendium

Dear Dr Cervas:

With my apologies, we received an additional reviewer's comments on your paper for you to consider as you start revisions. The reviewer agrees that your paper is acceptable with minor revisions, and just has some additional comments for you. Those comments are below, but please let me know if you have any questions at all.

Sincerely,
Eric Sampson
Statistics and Public Policy

This is an excellent and an important paper. But there are some issues, mainly with the presentation, which if addressed might improve the readability of the paper. They are being offered to the authors with no obligation that ANY of them be adopted. We believe the entirety of suggested changes should not take longer than an hour or two to incorporate and might make it easier to understand. I also have a few typos that I discovered.

p. 3 line 11 --- Republican members of Congress
p. 3 line 30 --- if they win in November --- this already happened
p. 4 line 6 --- ease of manipulating
p. 4 line 17 --- "like a hydra-headed monster" --- a little too colorful for me. Also cut "all of them"
p. 4 line 39 --- persist not persists
p. 6 footnote 6 --- This may be important enough to put in the text.
p. 6 footnote 8 --- cite it herein --- cut "to"
p. 7 lines 50-54 --- Isn't the point that Trump might have done much worse in the counties he lost in 2020 than in 2016?
p. 8 line 34 --- it becomes
p. 8 line 36 --- "red" country
p. 9 footnote 10 --- I am not exactly sure what is being said with "so the circle size itself is not a linear scale". Maybe the footnote is better dropped.
p. 10 Note: "Each county is shown as geographically …
p. 13 footnote 13. Line 33 Biden's percentage of the 438 electoral college votes in 2020 would only have gone up ---also delete "one"
p. 15 footnote 14 line 32 --- "The minority electorate increased in 2020 as compared to 2016 both in raw numbers and in proportion of the total."
p. 16 footnote 15 --- I don't think this footnote adds anything important.
p. 17. Line 30 --- "A further note of caution is that high statistical significance in regressions does not imply a causal relationship between variables. Either this is misplaced or it should be dropped. I would hope most readers would know this. But in any case it doesn't really belong in this section.
p.18 line 40 --- "and later vote share for TRUMP to be"
p. 19 line 5. "And about those facts and the statistical significance of the observed differences he

p. 19 line 5   "And about those facts and the statistical significance of the observed differences he is again quite correct." This seems redundant to me.

p. 19 footnote 19. I don't completely understand. Of course in states which are entirely one way or the other no comparison in possible.

p. 20 line 44 --- I would cut "Indeed, what would be surprising is if silly null hypotheses were NOT rejected."

p. 21 lines 51-56. I think it is confusing to add in the 366. Just say that "assuming a year of 365 days …."

p. 23 lines 3-24. I found this a little hard to follow. I guess you are saying conditional on a given name, it is …. Obviously the divisor is no longer 365.

p. 23 footnote 23 what is a self-weighted average? Also the frequency of names varies geographically.

p. 24 line 28 What is meant by data identifiers? Do you mean name and date and year of birth? How about characteristics?

p. 25 I don't think this section on Benford's Law is adding much. You have no example where the argument was made except in unnamed videos or tweets, not really things one needs to rebut. And the use of Benford's law, which is not really a law but sometimes an empirical finding, is rarely applied to anything but the first digit. I would drop this section.

p. 26 It is noted that the next two examples are more about human psychology and I am also inclined to drop these two sections for this reason. The first point is finding a little fraud is tantamount to finding a LOT Of fraud. Not much of an argument. Similarly I don't think the two footnotes 32 and 33 add much. However, I do think it is important to have the content of the initial paragraph starting on p. 26 on tip of the iceberg. As for straw man fallacy, I really don't think this adds much.

p. 28. Line 26 An example IS the empirically …

p. 35 lines 3-17. Unless I am misunderstanding something, you seem to be arguing that the argument against bell weather counties is that polarization makes them hard to find. You want to argue that once found, they don't work. That is different.

p. 36 line 3. "we could just as easily claim" delete "to"

p. 36 line 31. "claims that support levels in straight ticket votes and split ticket votes should be unrelated unless there is voter fraud." I don't understand what is being argued here. "Ayyadurai asserts that a negative correlation…" Negative correlation of what? I think we need a little math here to make it clear what is being said. Is this at some level of aggregation, like precincts or counties? I don't understand what the units of analysis are.

p. 37 line 26 "Republican vote share IN four Michigan …"

p. 38 line 21. "Now doing what Ayyadurai wishes us to require IS A simple correlation between r and either …"   But this wouldn't just be a positive or negative correlation. It would be +1 or -1. I guess he does this in the case where r(1) does not equal r(2). So maybe no change is needed.

p. 39 line 8 "the pattern of Trump drop-off is different." I am not sure what is meant by Trump drop-off.

p. 40 footnote 50 line 55. "they make is that which precincts are paired …"

p. 42 lines 20-37. I am a little lost. First, I am not sure what is meant by on average. Given that the mail and the in person votes sum to the total, conditioned on the total they have perfectly negatively correlated, in that a change from a mail to an in person vote moves one up one and one down one. Is that all that is being argued? And V(i) is a total vote on p. 42 and a percentage on p. 43 so a little rewriting would be helpful here.

p. 44. How many matched pairs were there. I am not sure what Lott was arguing from what is provided. Also, I think the mention of Simpson's paradox (actually the generalized version to regression, not the contingency table version) is speculative and maybe should be dropped. Lastly, I am a little concerned with the partisan nature of the argument in the sense that all of the errors are being made by Republicans. That is only to be expected since this was an election in which the democratic candidate won. But it would be nice to mention that this is something that both sides stoop to. I am assuming (hoping) there is a literature that mentions some democratic misbehavior. The authors are clearly well acquainted with this literature and if

there were a paper that made valid arguments of democrats misusing statistics in the past, that would be helpful to include either at the beginning or at the end just to avoid causing some hard feelings. If such a paper doesn't exist, you could just assert that you wouldn't be completely surprised to see similar arguments made by democrats in the future. Unless you think this is unprecedented, in which case forget this point.

Hope these comments are helpful.
Again, excellent paper.