Dear Professors Johnson, Mears, and Stewart (and copied coauthors),
There seem to be irregularities in the data and findings in five articles that you published together with two surveys. This document outlines those irregularities. I am requesting the data and analysis code, or at least the R/SAS/Stata/SPSS output from your analysis, so that I can attempt to identify the source of these errors. The first survey was conducted in 2008 and the second in 2013. The five articles, grouped according to survey, follow. After acceptance and online publication, but before print publication, Mears et al. (2019, p. 487) changed all of the tables in their paper because of a "coding error." The changes removed the standard deviations (fixing the mean-SD discrepancies), added variability to the standard errors, and placed zeros in the third decimal places of coefficients and standard errors. Because the earlier "online first" version of that article is perfectly consistent with the anomalies in the other four published articles, it is discussed here to demonstrate the pattern.

2008 Survey

- Johnson, Brian D., Eric A. Stewart, Justin Pickett, and Marc Gertz. (2011) Ethnic threat and social control: Examining public support for judicial use of ethnicity in punishment. *Criminology, 49*(2), 401-441.
- Stewart, Eric A., Ramiro Martinez, Jr., Eric P. Baumer, and Marc Gertz. (2015) The social context of latino threat and punitive Latino sentiment. *Social Problems, 62*(1), 69-92.

2013 Survey

- Mears, Daniel P., Eric A. Stewart, Patricia Y. Warren, Miltonette O. Craig, and Ashley N. Arnio. (2019) A legacy of lynchings: Perceived black criminal threat among whites. *Law & Society Review, 53*(2), 487-517.
- Stewart, Eric A., Brian D. Johnson, Patricia Y. Warren, Jordyn L. Rosario, and Cresean Hughes. (2019) The social context of criminal threat, victim race, and punitive black and latino sentiment. *Social Problems, 66*(2), 194-221.
- Stewart, Eric A., Daniel P. Mears, Patricia Y. Warren, Eric P. Baumer, and Ashley N. Arnio. (2018) Lynchings, racial threat, and whites' punitive views toward blacks. *Criminology, 56*(3), 455-480.

1) Anomalies in standard errors, coefficients, and p-values

In Stewart et al. (2018), tables 2-4 include regression results. The standard errors are identical to the third decimal place across models. The web links are to pictures of the tables. Stable standard errors are in yellow. Some stability in standard errors is normal, especially with a sample of this size, but this level of stability is unusual given the observed changes both in the regression coefficients and in the amount of explained variance. Combined there are 548 regression coefficients and standard errors in these three tables, but just one ends with a zero in the third decimal place. This is unusual because the distribution of third-decimal-place numbers (with rounding) should be close to uniform.

https://imgur.com/pcdB5XF

https://imgur.com/nEft7lQ

https://imgur.com/iDDs2as

The standard errors are also identical across models in Stewart et al. (2019). The following web links are to the tables. Stable standard errors are in yellow. None of the 348 regression coefficients and standard errors in these two tables end with a zero in the third decimal place, even though the distribution of third-decimal-place numbers (with rounding) should be nearly uniform.

https://imgur.com/I87pBjm

https://imgur.com/Vwbc50D

The same pattern occurred in tables 2-4 of the online-first version of Mears et al. (2019). The standard errors for the variables were identical to the third decimal place across the models. The following web links are to pictures of those tables, which were all changed between online publication and print publication. Stable standard errors are in yellow.

https://imgur.com/bJbzlh1

https://imgur.com/3fwiPr0

https://imgur.com/4lURlGr

The standard errors are also stable (to the third decimal place) across models in table 2 of Stewart et al.'s (2015) *Social Problems* article. The web link below is to the table. Stable standard errors are in yellow.

https://imgur.com/ls98UP5

The same kind of standard-error stability occurs in Johnson et al.'s (2011) table 2. The web links below are to pictures of the table, which has two panels. Stable standard errors are in yellow.

https://imgur.com/IryIT8Y

https://imgur.com/m5j3tr2

The distribution of p-values in these articles is also unusual. Across the five articles, there are 791 p-values, but not a single one falls between .045 and .105. This is highly unlikely because p-values are uniformly distributed under the null hypothesis, so one would expect numerous p-values in this range due to chance. For example, in the next six-point range, from .105 to .165, there are 35 p-values. In the six-point range after that, from .165 to .225, there are 19 p-values. And in the six-point range after that, from .225 to .285, there are 27 p-values. Actually, none of the other six-point ranges above .045 is empty, except the .045 to .105 range. This web link is to a figure that shows this.

https://imgur.com/azGDAla

2) Discordant means and standard deviations for binary variables

For binary variables, it is possible to calculate the standard deviation knowing the sample size (N) and variable mean (P), the proportion of respondents coded 1. The formula is: SD = sqrt {[P*(1 – P)*N] / N – 1}. In both Stewart et al. (2018) and the online-first version of Mears et al. (2019), the standard deviations for four of the binary variables are wildly inaccurate, too inaccurate to be due to rounding. Mears et al. (2019) changed this table before print publication. The web links are to the Stewart et al. (2018) and original Mears et al. (2019) tables with the errors highlighted. Given the listed means and sample size, the correct standard deviations are: married (.499), education (.497), political conservative (.495), owns home (.463).

https://imgur.com/rOzyqDF

https://imgur.com/5lS6vc

In Stewart et al. (2019), the standard deviations are wrong for nine binary variables. The web link is to the table with the errors highlighted. Given the listed means and sample size, the correct standard deviations are: married (.50), education level (.50), political conservative (.50), owns home (.46), Southwest (.40), Northeast (.33), Midwest (.38), West (.36), and South (.50).

https://imgur.com/9dUcMUs

In Stewart et al. (2015), the standard deviations are wrong for six binary variables. The web link is to the table with the errors highlighted. Given the listed means and sample size, the correct standard deviations are: married (.49), education level (.49), political conservative (.50), owns home (.41), Southwest (.38), and South (.50).

https://imgur.com/ShYo61i

The Johnson et al. (2011) article includes nine binary variables where the means and standard deviations do not match. The web link below is to the respective table with the errors highlighted. Given the listed means and sample size, the correct standard deviations are: white (.35), black (.30), Hispanic (.20), married (.49), education level (.49), political conservative (.50), owns home (.41), Southwest (.38), and South (.50).

https://imgur.com/8yxseEK

3) Identical descriptive statistics across different samples

Johnson et al. (2011:411) use a mixed-race sample in their analysis, whereas Stewart et al. (2015: 76) use a sample of non-Hispanic whites. Although these two samples have different racial compositions, they have identical means and standard deviations (to the decimal places) on twenty of the variables. It is unlikely that this is the result of the authors accidently including the wrong table in one of the papers, because the samples do differ on several variables (e.g., county ethnic and racial composition and concentrated disadvantage). The web link is to the tables. Matching descriptive statistics are in yellow. Sample racial characteristics, which are only in Johnson et al. (2011), are in red.

https://imgur.com/lbkV02U

4) Identical descriptive statistics across samples of different sizes

Stewart et al. (2018) use a sample of 1,144 Southern whites in 90 counties. Stewart et al. (2019) use a sample of 2,408 Whites in 168 counties. Despite the large difference in sample size, these two samples have identical means and/or standard deviations on ten variables. It is unlikely that this is the result of the authors accidently including the wrong table in one of the papers, because the samples do differ on several variables (e.g., family income, political conservatism). The following web link is to the tables. Matching descriptive statistics are in yellow. Sample size differences are in red.

https://imgur.com/pW9YvrX

5) Unusual changes in sample size over time

Johnson et al. (2011) use data from a survey conducted by a polling firm called the "Research Network" in 2008. The same data are used in Stewart et al. (2015). Although the survey was conducted in 2008, the total sample size grew from N = 1,184 in Johnson et al. (2011) to N = 1,379 in Stewart et al (2015). Yet, the survey particulars remained unchanged (e.g., 54.8% response rate). And many of the descriptive statistics stayed the same (e.g., in both samples, the mean age is 47.12 and mean family income is $62,700).

6) Incorrect statistics and distributions

In the *Social Problems* article by Stewart et al. (2015, p. 76), the authors wrote: "The breakdown for annual household income was as follows: about 25 percent of the sample reported earning less than $50,000; around 14 percent of the respondents earned between $50,000 and $75,000; 9 percent of participants earned between $75,000 and $100,000; and about 12 percent of the sample reported earning more than $100,000. The median family income in the sample is $40,900, with a mean of $62,700." There are two problems with this. First, the percentages do not add up to 100% (25% + 14% + 9% + 12% = 60%). Second, the median income cannot be $40,900 if only 25% of the sample earned less than $50,000. By definition, the median is the 50% mark.

https://imgur.com/DegoLVt

In Johnson et al. (2011, p. 412), the authors wrote: "The breakdown for annual household income was as follows: Approximately, 52.4 percent of the sample reported earning less than $50,000; approximately 21.7 percent of the respondents earned between $50,000 and $75,000; 13.9 percent of the participants earned between $75,000 and $100,000; and approximately 12.0 percent of the sample reported earning more than $100,000. The median family income in the sample is $40,900, with a mean of $62,700." If 52.4% of respondents had family incomes under $50,000 and the median family income was $40,900, then only 2.4% of respondents had a family income between $40,900 and $50,000. This is very unlikely, given the high prevalence of family incomes in this income bracket in the US population. Furthermore, given the large differences in income distributions between this paper and the 2015 *Social Problems* article (52.4% under $50K vs. 25% under $50K), it is odd that both samples have the same median ($40,900) and mean ($62,700) family incomes.

https://imgur.com/7mE8lzD

In Stewart et al. (2018, p. 471), the authors provide the only exact p-values in the study for supplementary analyses: "We thus estimated models consistent with those in tables 2 and 3 but focused instead on the Black respondents (n = 200). Black lynchings did not yield a statistically significant effect on punitive-Black sentiment (b = .065, standard error [SE] =.063, p = .35) or the Black-White punitive sentiment ratio (b = .059, SE = .058, p = .39). But the provided p-values do not match the provided coefficients and standard errors. The values should instead be: p = .30 and p = .31.

https://imgur.com/X65TQkQ

7) Unlikely survey design and data structure

In Johnson et al. (2011, p. 411-413) and Stewart et al. (2015, p.76) the authors wrote that the Research Network administered the 2008 survey using "a two-stage modified Mitofsky–Waksberg sampling design" to randomly sample "American households with either landlines or cellular Phones." This yielded a high degree of clustering, with more than ten respondents in each county, on average. But Mitofsky–Waksberg sampling is rarely used for cell phones. Additionally, in neither study do the authors discuss how they handled errors in matching wireless numbers to counties. According to a Pew Report by Christian et al. (2009): "The geographic information derived from cell phone numbers is subject to a great deal of error … the sample and zip code-derived county do not match for nearly four-in-ten cell respondents (39%)."

.
http://www.pewresearch.org/2009/07/09/accurately-locating-where-wireless-respondents-live-requires-more-than-a-phone-number/

The 2013 survey also used "a two-stage modified Mitofsky–Waksberg sampling design" to randomly sample "households with landlines or cell phones" (Stewart et al., 2018, p. 460). Although 2013 and 2008 surveys have identical designs, both unusual, it does not appear that the Research Network administered the 2013 survey. The Research Network is not mentioned in any of the three articles that use the 2013 data.

The differences between the 2013 sample characteristics and the US population are also difficult to reconcile with the use of random sampling for a sample this large. The total sample includes 2,736 Americans, of whom 2,408 are "non-Latino White respondents (N = 2,408)" (Stewart et al., 2018, p. 461). This means 88% of respondents are non-Latino Whites. According to the US Census, however, only 60.7% of Americans are non-Latino Whites. A random sample of this size should not be 27 percentage points off the population value. Stewart et al. (2018, p. 461) write that the "southern sample [of Whites] in our analysis consists of 1,441 respondents who resided in 90 counties across these 11 states." This means that 60% of all whites in the sample (1,441/2,408) lived in 11 southern states. This is odd, because in their *Social Problems* article, Stewart et al. (2019) give a different percentage:

https://imgur.com/tP8EWhk

It is also improbable because, according to the US Census, only 37.5% of all Americans lived in the South in 2013. Moreover, the percentage of White Americans living in the South was even lower, because the South is the second most racially diverse region. In a large nationally representative survey, one would expect about 30-35% of White respondents to live in the South, not 55-60%. A large random sample should not be off by over 20 percentage points.

· https://www.census.gov/popclock/print.php?component=growth&image=//www.census.gov/popclock/share/images/growth_1530403200.png
· https://statisticalatlas.com/United-States/Race-and-Ethnicity#figure/region

None of the articles using the 2013 survey list a funding agency or grant number, which is surprising, because a nationally representative, dual-frame, telephone survey of 2,736 Americans would cost well over $100,000.

· https://www.surveypractice.org/article/3168-the-changing-costs-of-random-digital-dial-cell-phone-and-landline-interviewing

Respectfully,
John Smith