Gmail by Google

Jim Hopper <jim.hopper10@gmail.com>

question

Jackie White <jwwhite@uncg.edu> Fri, Sep 4, 2015 at 10:39 AM
To: Jim Hopper <jim.hopper10@gmail.com>
Cc: Kevin Swartout <kswartout@gsu.edu>, David Lisak <david@davidlisak.com>

Jim, I plan to double check the public dataset and upload new information as necessary. In the meantime, you will have to be patient. I am retired and no longer have access to SPSS unless I physically go to campus, which I rarely do. As I spend quite a bit out of town visiting my adult children, I will have to find a time I can set aside to work on this.
JWW

Sent from my iPad

On Sep 3, 2015, at 8:20 PM, Jim Hopper <jim.hopper10@gmail.com> wrote:

> Hi Jackie,
>
> Thanks for getting back to me. I'm not sure what you mean by "a few coding errors."
>
> We brought to Kevin's attention that there were (1) about 110 erroneous missing values for the sophomore year assessment and about 70 erroneous missing values for the junior year assessment, and (2) a few other discrepancies with respect to missing data between the dataset of dichotomous "R/rape" variables (that he used for the JAMA paper and sent to us) and the X____ variables available in the public database (which have frequency information and are not effected by the command that recoded "missing" and "no response" into "never" to create the dichotomous variables on which Kevin's JAMA-paper analyses were based).
>
> Kevin never provided any information about the "no response" = "never" and "missing" = "never" recode command for the dichotomous SES item variables (which is on p.512 in the PDF of the codebook and clearly antedates his involvement with the data). He said he would get back to us about that, but never did.
>
> (On the 110 and 70 erroneous missing values, Kevin said this: "the missing data were created (unsystematically, as you noted) when Dr. White's data were restructured (wide-to-long format) for MLM analyses. I re-constructed the dataset from the original frequency data and re-ran the derivation model; the results are consistent with those published in the paper, and the interpretations and conclusions are unaffected. We have alerted the journal editor to this issue.")
>
> So the reason I'm writing to you is because Kevin appears to have decided not to answer the question about the ("no response" = "never" and "missing" = "never") recode command, which he said he would answer in a prior email. (Indeed, he has not answered many of the questions I've asked him, perhaps in part because he feels I've not been respectful in my emails, as suggested by a message he sent this morning to Michael Kimmel this morning.)
>
> And because that recode command antedates his involvement in the project, it probably made sense all along for you to be the one to answer that question (or to try to answer it-- and I do hate to ask but hope you're willing to try, despite your retired status).
>
> Also, it's important for you (and Kevin) to know that our goal is not merely to re-conduct the analyses that Kevin conducted for the JAMA paper. Much more important, we want to understand -- and to help everyone who is interested in campus sexual assault understand -- how the dichotomous data Kevin used for the JAMA paper are related to the raw frequency data contained in the public database (including how many rapes of different kinds, at different assessments, were

self-reported by subjects assigned to specific latent classes).

This is because our analyses of the public data indicate that many of the subjects that Kevin has defined as not serial rapists actually reported committing multiple rapes at particular time points, sometimes more than 5. For example, we believe it is important for other scientists, activists, policy makers and the media to know just how many rapes those subjects Kevin has defined as not serial rapists have actually reported committing, and what percentage of rapes are committed by those who report committing multiple rather than just one rape (overall and at particular time points). The use of the dichotomous variables and the nature of the latent class analyses used in the JAMA paper have the effect (intentional or not) of obscuring that very important information.

So that's why I'm asking you about that recode command again. We want to really understand the sexual assault perpetration data, inside and out, and make sure our analyses are totally grounded in as valid as possible data and totally transparent to everyone. That includes understanding the relationship between the dichotomous SES data and the corresponding raw frequency SES data. That includes understanding the relationship between the original "derivation" dataset (with all of those erroneous missing values) used for the JAMA analyses and the raw frequency data (which is why I've been quite frustrated that Kevin hasn't shared that dataset with the subject ID #s and the R variables); etc.

Does this make sense? Can you help us with that recode command and impacts it had on the data? (My attempts to figure it out suggest that only one subject may have been affected, but I just can't be sure, since the original variable that was recoded [which presumably had values of "no response" that were distinct from "missing", otherwise the recode command would not have been created] is not to be found in the database.)

There may be other questions we have too, about the public database. For example, for the junior year assessment, a whopping 154 subjects have "don't know" for the value on the item about their sexual orientation (SXORIEN4); that's 34% of the 457 subjects who still remained that year. This could mean the validity of the data was seriously compromised at that point, with subjects making random responses to items and/or providing deliberately misleading responses. Or something could have gone wrong with data entry for that item that year.

Seeing all those missing values for sexual orientation junior year (and therefore concerned about data integrity and validity), I started looking at other demographic variables for each assessment period, and found:

> For freshman year assessment, there are 452 missing values for the “relationship status” variable, and 602 missing values for “marital status,” but only 211 missing values for XFORCSI2 – those are differentials of 341 and nearly 400 missing values.
>
> For sophomore year, there are 754 missing values for “college educational status” and 395 for “relationship status” (no marital status item that year), and 398 for XFORCSI3.
>
> For junior year, 738 missing for “college educational status” and 551 for “relationship status,” and 552 for FORCSI3.

These findings raise some questions: Why so many missing values for “college educational status”? Why so many (more) missing values for relationship status in freshman year but not sophomore and junior years? What if anything does this tell us about data integrity overall?

Do you recall noticing these issues back when you were immersed in the data years ago, when you were evaluating the quality of the dataset, cleaning up mistakes before testing hypotheses, etc.?

Again, I'm so sorry to burden you with this. And I'm really sorry if I'm not being as tactful as I could be or you would like. (Please tell me if you feel I'm being disrespectful in anything I'm writing or have written to you.)

But once Kevin effectively threw down the gauntlet against David and "the serial perpetrator hypothesis" -- with everything he did to create, publish and promote the claims he makes in the

JAMA paper -- he created the conditions that have compelled us, by necessity, to determine the bases of his research and claims, including the "derivation" dataset that happens to belong to you.

Also, because email can lead to misunderstanding and potentially conflict-causing (or -escalating) interpretations of others' actions and motives, and because I'm not always as polished or tactful as I could be, it might be very helpful if you and I had a phone conversation.

Best wishes,

Jim

On Thu, Sep 3, 2015 at 9:52 AM, Jacquelyn White <jwwhite@uncg.edu> wrote:

> Jim, I have been out of town at a conference and visiting family. Now I am playing catch up with all my emails. I will be working on getting the original data set cleaned up, but will take awhile.Remember, being retired is making it harder for me to deal with all this.
>
> Kevin did work through everything and determine that a few coding errors happened, but it was all random. I thought he shared that with you.
>
> jackie
>
> On Tue, Sep 1, 2015 at 10:50 PM, Jim Hopper <jim.hopper10@gmail.com> wrote:
>
> > Dear Jackie,
> >
> > Are you willing to answer questions about your male 1990-1995 dataset -- if the questions are not about the analyses Kevin conducted or the specific dataset he used for the JAMA Pediatrics paper, which he has been unwilling to share with us (that is, the dataset with the five R/rape variables [which he gave us] and the subject ID numbers that can be matched to the public dataset [we he has been unwilling to provide])?
> >
> > There are still some things that we're trying to understand about the publicly available dataset.
> >
> > For example, we're still wondering about the recode command that recoded "missing" and "no response" values into "never [committed the act]" values. Specifically, as I mentioned before, we're wondering why such recoding was done and what, if any, effects it may have had on the validity of some variables in the public dataset.
> >
> > And a few other questions have come up as we've dug deeper into the full public dataset.
> >
> > I'm sorry if this is putting you in an awkward position. Thanks for your consideration.
> >
> > Regards,
> >
> > Jim
>
> --
> Jacquelyn W. White, PhD, Emerita Professor of Psychology
> Senior Research Scientist
> Center for Women's Health & Wellness
> 126D HHP
> School of Health and Human Sciences
> PO Box 26170 (for regular mail)
> University of North Carolina at Greensboro
> Greensboro, NC 27402-6170

1408 Walker Avenue (For shipping)
401 HHP
Greensboro, NC 27412
Fax: 336- 334-3238
jwwhite@uncg.edu

336-337-2100

https://sites.google.com/a/uncg.edu/jackiewhite/home

Co-Chair, National Partnership to End Interpersonal Violence (www.npeiv.org)