Proxies

Youtube generated subtitles

so far we have talked about why we should strive for a mixed methods approach this incorporation of qualitative and quantitative methods and why we need to consider multiple ways of measuring those things that we are interested in particularly when they are more abstract Concepts things like engagement intelligence maybe even success

we've even sort of spoiled The End by saying why we care about this diversification of methodology and measurement we said it all comes down to finding that consistency among approaches that allows us to be more confident in our conclusions more confident in our ability to generalize the results outside of sort of the parametric space that we built our models in

right now I would like to go back to the example of using attendance to represent student engagements okay if you if you review what I said um I said that attendance was a manifestation of student engagement or put another way student engagement causes that's what the arrow is causes or influences student attendance now is that the only cause of attendance rates no right there are very few things particularly at higher levels of complexity like people that have singular causes you can't just do one thing okay but it is a cause it is an influence of attendance What If instead of attendance I'm interested in cognitive abilities I'm hiring someone okay could I use something like the number of years educated as an approximation

folks who are in school longer tend to have higher cognitive abilities so seems like a reasonable approximation sure I could do this right but we do need to take a note of this unlike in the case of student engagement where I said student engagement leads to higher rates of attendance we don't say that cognitive we don't say that cognitive and cognitive abilities lead to more years educated okay in fact for the sake of all the Educators in the class I hope that years educated can sometimes go the other direction and increase cognitive abilities otherwise what are we doing um educating folks um we know that there's an association between the two but we don't necessarily know the direction we just know that we can use the number of years educated as a stand-in or a proxy for this thing called cognitive abilities

now we see examples of proxies used all the time in various Industries okay you have things like the BMI as a proxy for body fat percentage it gets used in medicine it's used in Insurance Actuarial tables all that good stuff you have per capita GDP or GDP per capita this often gets used by economists and sociologists as a uh as an as a proxy for the quality of life okay um SAT scores right SAT scores are a proxy for this thing called Scholastic aptitude or Scholastic achievement right but it is in fact just a test that is meant to try to um to ascertain it okay

but none of these variables none of these proxies are in fact the thing that they are representing they're not even a

manifestation of those things that they are representing they are a stand-in or an approximation of these more complicated these more complicated constructs okay we know that they are associated but they're not likely to be caused

now the use of these proxies is fine in fact when it comes to things like prediction if I'm building a series of predictive models proxies are great because I can use one variable instead of maybe having to use 10 or 12 or 13. okay if I can just use per capita GDP and have that be an estimate for the quality of life instead of having to go through take a series of different metrics and combine them all into the model I end up with a more parsimonious a more parsimonious model and something that's a lot easier for me to interpret okay

there is nothing inherently wrong with using a proxy as long as everyone remembers what the proxy is it is a stand-in for a variable not the variable that it represents Walsh cats and Seacrest is 2002 is a paper I'm going to talk about here in a little bit in it they point out that a lot of the way that research gets done their example was in medical research but this applies elsewhere as well it's sort of automated in that whenever you run a particular analysis you have a set of demographic variables that you throw in sort of without thinking age sex location so to speak now this process can it's it's helpful this process is helpful because it can lead researchers to find that ever elusive p-value to publish their paper right it allows them to find differences between groups that they might not have looked at have they not Incorporated those demographic characteristics but when they find those differences and don't explain why such differences might exist or they use that opportunity to really just rehash um stereotypes about particular groups it's not helpful in fact it can be actively harmful so it seems like in many cases the demographics that folks sort of just automatically include in their studies they're really just treated as covariates right predictors that they just don't really care about or don't really like variables that they're going to control for but not really interesting to the hypotheses or the effects that they're actually trying to find okay but when you stop and you try to unpack what something like sex or race can speak it is a stand-in for you can learn something more about the mechanism of what's going on underneath and that can lead you to a potential Target for something like an intervention

uh in their study walls cats and Seacrest uh 20 uh 2002 um they constructed they they created a construct okay and that construct represented available resources for disease management um pertaining to folks who had uh type 2 diabetes uh veterans who had type 2 diabetes okay um it in court it included things like um economic marginality domestic and family workload domestic help positive family relations salience of religion proactive response to illness negative impact of illness social support and social cost okay essentially they created this construct that was able to say something about the various um the the various social and emotional and ecological resources that were available to folks who were struggling with managing their type 2 diabetes okay they went ahead they created this composite and they put they pitted it up against uh self-reported ethnicity something that has been found to correlate with various outcomes okay

they compare to the explanatory power of this composite to self-reported ethnicity and found that not only not only did this composite explain more of the outcomes in every instance than ethnicity did

but it didn't actually leave a whole lot for ethnicity to account for so they ended up creating this composite that was able to account for more uh more of these outcomes like adherence health-related quality of life disease management and utilization all of these were statistically significant except for adherence um they were able to account for more of these outcomes with their their composite of this sort of adaptation to their disease so much so that it was able to eat up all most of nearly all of the variants that ethnicity was explaining for meaning that the differences seen in ethnicity seem to be driven by this this adaptation factor that they created

so the whiplash why did I go from the last two lessons telling you that you should try to use several different methods and several different measures in your research whenever you can and now I'm saying using this obviously reductive procedure of proxy variables is okay sometimes well I did it to demonstrate something it's something I've said before but it's something that I will probably say several more times there is not a single recommendation or a set of rules or prescriptions that I or anyone else can give you to determine how you should handle your data analyzes and your research if you're still hanging on to the idea of there being a best then you should really disavow yourself of that like yesterday okay I want to read um I want to read a quote from uh from unobtrusive measures um it's in chapter seven uh it's it's kind of a long quote and I'm sorry but I'm gonna go ahead and read it it's gonna be uh in two two slides it actually starts with a quote within a quote um by historian of psychology EG boring

as long as a new construct has only a single operational definition that it received at Birth it is a construct when it gets too alternative operational definitions it is beginning to be validated when the defining operations because because of proven correlations are many then it has become then it becomes reified

this means obviously that the notion of a single critical experiment is erroneous there must be a series of linked critical experiments each testing a different outcropping of the hypothesis it is through triangulation of data procured from different measurement classes that the investigator can most effectively strip plausibility of rival explanations for his comparison and here's the big one

the usual procedural question asked is which of the several available data collection methods will be best for my research problem we suggest the alternative question which set of methods will be best with best defined as a series which provides data to test the most significant threats to a comparison with a reasonable expenditure of resources

in short think carefully about what methods and measures you are going to use think about what the data analytic Journey has

been and where it might be going where did you come from where did you go

something something Cotton Eye Joe

there is nothing in statistics or methodology that requires you to do something stupid do what works for the data set that you have and the analysis that you have trying to answer the questions that you have

we will continue our discussion on measurements in the next uh in the next uh couple of weeks okay we will probably start with unpacking this idea of reliability and validity a little bit further and I hope that this has sort of been a good anchoring for you um in terms of sort of the overall perspective that you want to keep I want you to kind of keep in the back of your head uh going forward