Southern New Hampshire University

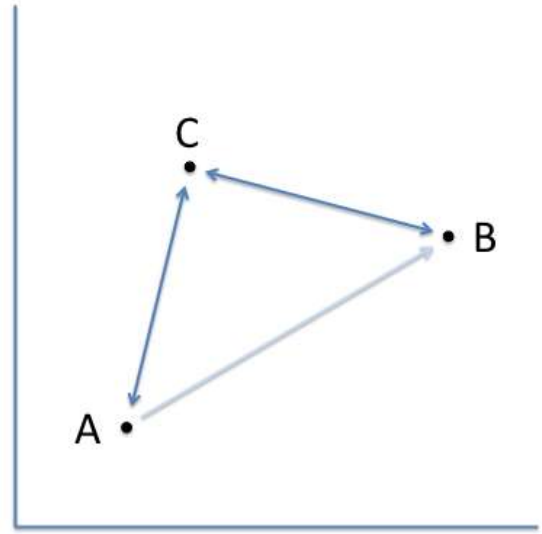
DAT 520 Module Nine Overview

Ethics in Decision Analysis

We have been considering ethics in decision analysis since the early modules, especially regarding the von Winterfeldt paper and in the discussions about how best to apply decision modeling techniques to various scenarios. In this module, we will look specifically at the intersection of ethics and medicine, with an eye toward decision making in business and other realms. How can something be technically correct, in terms of being fed by valid information, but be morally corrupt? It is also possible that all the available information is not even the correct information. So what can you do to ensure that your research is both morally applicable and meets your personal ethical criteria? What are some things you can consider in your writing to address these topics? One way might be to consider the downstream effects of a decision made one way versus another. If you put the highway through one neighborhood versus another, what are the ramifications? What about if your company purchases a giant shipment of Software A versus Software B? What are the financial impacts and the cultural impacts on the worker community? These are but a few of a number of ways to consider ethics in decisions.

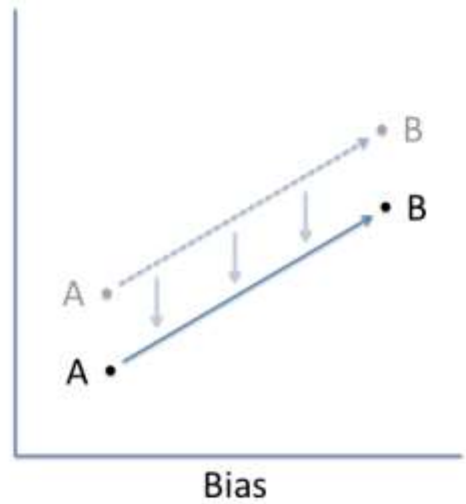
Bias, Error, and Confounding in Decision Analysis

No statistical report or lesson would be complete without some mention of confounding and bias. They are constant threats in the background. It is very important to consider the role of errors that can creep into your work. A thorough discussion of the weaknesses of any analysis is a critical piece of your report because research is additive. Each new study you embark on can be made stronger than the last one by appreciating what you can do better. Two main ways to interpret weaknesses of an analysis are taking on the subjects of confounding and bias.



Confounding

The word confounding means “mixed up” or “confused.” Confounding is best explained situationally. In scientific usage, it means there is a third variable, *C*, which gets in the way of observing the effect between *A* and *B*. So, you may be thinking that *A* influences *B* in a certain way, but really, it is *C* that is getting in there and causing you to see a relationship where one may or may not exist, and it is cloudy (or mixed up) at best. The confounder *C* is related to both *A* and *B* independently, thus the two-headed arrows. But its presence masks the true relationship between *A* to *B*. You may be aware of confounding, or you may not be, but its possibility is ever-present in any scientific pursuit and needs to be considered thoroughly. In terms of decision analysis modeling, confounding might mean you have selected relationships between two variables in your model, but there could be a third variable in between that relationship, causing the results you are seeing. For example, if you have a model that directly relates age to type of car purchased, there could be a third variable in between age and car model, such as income, since older people generally tend to have bigger incomes than younger people, therefore confounding the types of cars different age groups tend to purchase. You should try to actively consider these possibilities when constructing your models.



Bias occurs when there is a systematic error in measurement or analysis. You may think you are measuring the relationship from *A* to *B*, in bold text on the diagram, but the true relationship of *A* to *B* is masked, as indicated by the shaded *A*→*B* on the diagram. Because of a systematic error in collecting the data, the true nature of the data has been obscured, leading to an underestimate in this case, indicated by the three short down-arrows.

Biases come in a number of different forms. Some are errors in analysis, misapplications of analysis, or are the net result of larger processes, such as a systematic exclusion or inclusion of certain types of research due to selective funding or publishing. For example, **publication bias** exists because medical journals tend to publish clinical trials with positive results, rather than trials that show no effect or a negative effect. Over time, the result of publication bias is an overestimate of the usefulness of drugs and treatments, because the favorable ones get the most airtime.

In terms of decision analysis, bias might relate to how the data was collected, who collected it, for what purpose, or how the data was processed after collection. You might also consider what variables were not considered, which is also known as **omitted variable bias**. You may not know the answers to any or all of these possibilities, but you can speculate. To summarize bias for decision analysis, remember the phrase "consider your sources."

Error typically refers to statistical error. In terms of discrete probability analysis, we have not taken up error estimation formally in this class. However, for our purposes, we might consider qualitatively if we have enough data to draw the right conclusions. For example, if you only have five data points and are making a probability based on those five data points, how great is the possibility of error creeping into that analysis? Probably very high. But if you have thousands of collected data points and are converting those into a probabilistic system, the likely error rate will be much lower. For decision analysis, even if you are unable to quantify the amount of error, you should be at least thinking about error qualitatively and how to describe its presence.

Internal and External Validity

Another way to characterize limitations of your model is to spend some time thinking about external and internal validity. **External validity** refers to the context of your data and decision analysis model in the larger world, such as where it came from, the intended purpose of the data, who collected the data and how, as well as understanding any confidentiality or security issues for that data. Also consider who is making the decision or reading the report. Consider if there are possibly better alternative data sources for the same information. Consider if there are alternative models that may serve different audiences. All of these things get at the external validity of your decision model.

Internal validity refers to the actual values and character of the variables and how they interrelate within the data set. With what you know about the data, are the values plausible within the data's external context? What, if anything, will you do about wild and outlier values? Can the number of observations in the data set be considered complete, or will more be needed? If you are using two or more data sets, is it reasonable to relate them to one another within the context of your decision analysis? All of these aspects get at the internal validity of your decision model, considering things like internal consistency of your data, whether different data sets play nicely with one another, whether it is reasonable to be using specific variables the way you are using them.

Why are we so worried about the limitations of a model? We do it because all of this thinking about how the current research is limited may lead to future research that transgresses the current limitations. By appreciating those gaps in knowledge upfront, the current research might point the way so that we know where to go next.