Project MOSAIC (mosaic-web.org)

Modeling Misconceptions

Feb 10, 2012 M-Cast

Remember: These things are being written as mis-conceptions. They are meant to be wrong!

- Misconceptions are deep-seated misunderstandings about concepts
- Misconceptions can arise from being in class or can be brought to class (prior knowledge)
-

Post the working concept inventory document. It will be posted on the MOSAIC site (password accessible, so that only instructors can browse it).

Put in links in this document for contact information.

Initial misconception in statistics: Students think that models are right.
There's only one model for a given phenomenon.
Need to have a way to evaluate how "good" a model is. There is no final model.
Definition of a model: A representation for a purpose. How to Model It --- great book!

Once a model has been created, it can be used for any purpose, even if that's not the purpose that the model was developed for. [[Example: Group-level versus individual-level models in statistics. Can't say a group-level model to say something about individuals.]]

[[Example: Reports from the model. Models of means are applied to individual people. Use of p-values to indicate strength of effect.

Cycle of modeling. You don't start with a correct model. You need to start somewhere, even though it might be wrong, and then interpret and evaluate it to see the extent to which it should be changed. In statistics: examining the residuals to see if there's something that remains to be explained, e.g., structure in the residuals.

Students take too seriously the limited set of models that they see in class. They think that these are the only possible models.

Watch out for models whose purpose is pedagogy, rather than illustrating real modeling practices. EXAMPLE: Use of cubic model in calculus when talking about optimization.

Example of a teaching context that shows the evolving nature of models:
1) in statistics add in new model terms, extending a model
2) in physics/engineering, use of Buckingham Pi. Propose the physical properties that you think play a role, construct the dimensionless groups, collect data to see if those dimensionless groups do the job or if something is missing. Example: Pendulum period: does air density play a role.

Purpose of Modeling

1. You model the "thing", not the "thing" in a context of a decision or problem.

2. The purpose is subjective, and subjectivity should be discouraged.
3. A correct model is correct for all purposes. Dan Flath: the "Don Quixote disease", to believe that your model is the truth, forgetting that it is only a model, and using it in inappropriate ways.
4. Mathematical models are about math, not the context. Dan Flath describes this misconception as the belief that "good math models can be produced with limited knowledge of the application context, if you know enough math. We actually encourage this in math classes by limiting the class time we spend on teaching the science of the application. In other words, our topics are almost always math centered rather than application centered."
5. The purpose is to translate reality into mathematical terms.

Architecture

1. Models must be based on first principles as in Newton's Laws.
2. Models must look like the stuff you learned in algebra: equations and formulas that you can write with a paper and pencil.
3. You can't model a probability; probabilities are uncertain.
4. It's always *x* and *y* in the sense that $y = f(x)$ *--- So, use different variables when you teach. What's the concept here? That there's only one explanatory variable possible.*
5. *Means and proportions. Maybe more about regression to demonstrate that what you're modeling is variation rather than the central value. We put too much emphasis on the fitted part, rather than the residual.*
6. The architecture of the model doesn't determine the sorts of solution techniques (you can always use a given set of solution techniques, e.g. analytic, whatever the structure).
7. If you adopt a mathematical structure, reality will follow. Example: Fit a logistic curve to cumulative oil discoveries and see if we are past the inflection point.
8. There can only be one variable doing the explaining.
9. There is always some pure mathematical form for a model.
10. There is "one right/best model"
11. Units don't matter. In reality, we do need to think about dimensions and units.

Operationalization

1. This doesn't involve choices.

Interpretation

1. Models are right or wrong in the sense that arithmetic or algebra is right or wrong.
2. Complexity tells you if the model is right.
    a. More complex means you're getting closer to reality. (Models must be realistic.) EXAMPLE: In the social sciences, since they work with observational data, toss in every covariate in the kitchen sink, instead of giving thought to what might be playing a role. Econometrics using "economic theory" criterion to determine if a model makes sense, regardless of what the data say. IF THE DATA and theory don't match, people sometimes assume that the data is wrong.

    b. Simpler is always better. (Occam's razor) "Absence of evidence is evidence of absence" --- not finding something means it's not there. EXAMPLE: Modeling the birth rate of boys and girls based on China's one-son policy. What would happen if families keep having children until they have a boy. Students take at face value the assumption that the male/female ratio is 50:50. See number 8 below on sensitivity.

3. Error bars are optional. (Similar to 8.)
4. Statistical significance means importance.
5. You can't evaluate degrees of approximation. Either it's on target or not.
6. That the model and reality should be the same.
7. If you've reproduced a phenomenon with your model, you've explained the phenomenon. Chad Topaz: Showing that you can model swarming behavior by a differential equation that has things move in a circle doesn't mean that you've explained the behavior.
8. *F=ma* isn't a model, it's reality!
9. No need to worry about how the model results depend on "small" changes in the parameters. "Sensitivity"
10. No need to worry about how the model results depend the detailed structure of the model. "Robustness"
11. "Is this model right?" is a sensible question. Chad Topaz: "How do we get students to stop asking if the model is right and instead ask, `Are the model results useful?'"
12. A better fitting-model is always better.
13. Group-level models say something about each individual.
14. Confusion of causation and association (e.g. a strong relationship infers causation)
15. After a model has been "built", it works for any purpose (regardless of the modelers' initial purpose for building it)