

Model evaluation

Jonathan Dushoff, McMaster University

<http://lalashan.mcmaster.ca/DushoffLab>

DAIDD 2015

<http://www.ici3d.org/daidd/>

Do I have a good model?

- ▶ What is my model trying to accomplish?
 - ▶ Evaluating plausibility;
 - ▶ Generating hypotheses
 - ▶ Prediction
 - ▶ Extrapolation
 - ▶ Mechanistic understanding

Statistical philosophy



OBEY^{THE} Kitties
or else...

Outline

Conceptual models

Prediction

Model Validation

Model Evaluation

- Goodness of fit

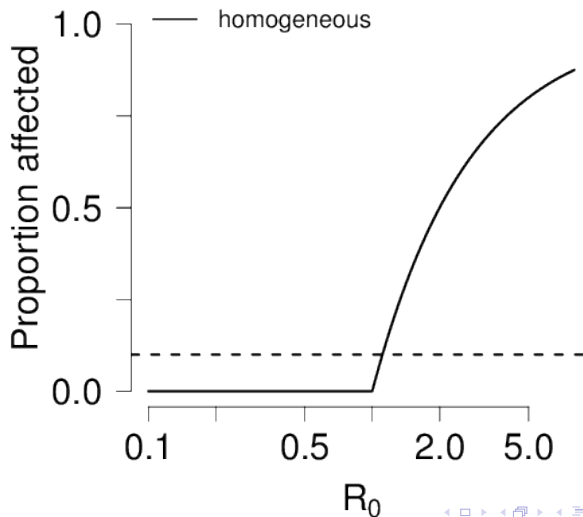
- Capturing patterns

- Going beyond

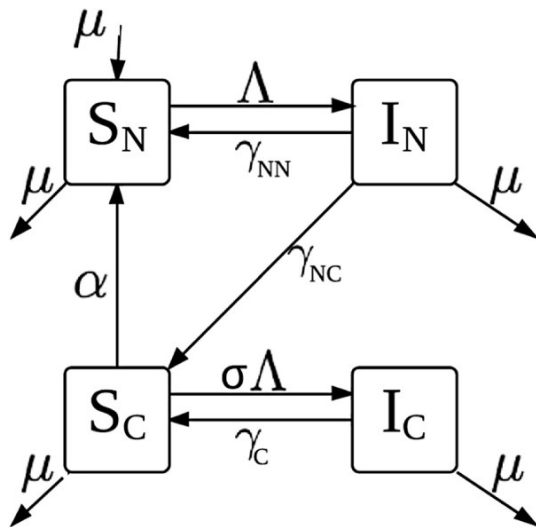
Conclusion

Disease thresholds

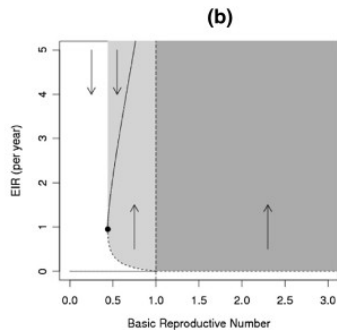
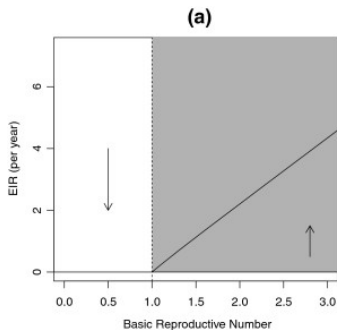
endemic equilibrium



Effects of clinical immunity



Bistability



Outline

Conceptual models

Prediction

Model Validation

Model Evaluation

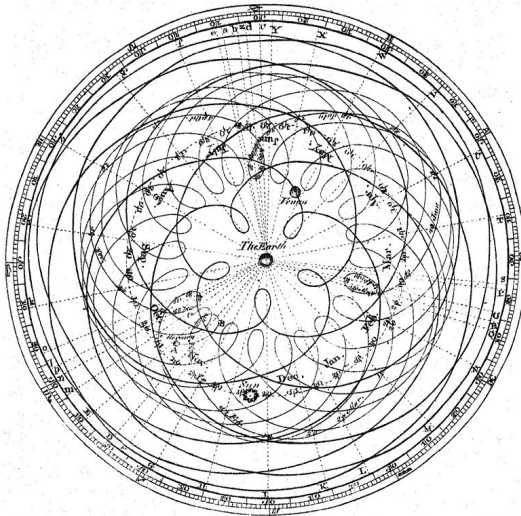
- Goodness of fit

- Capturing patterns

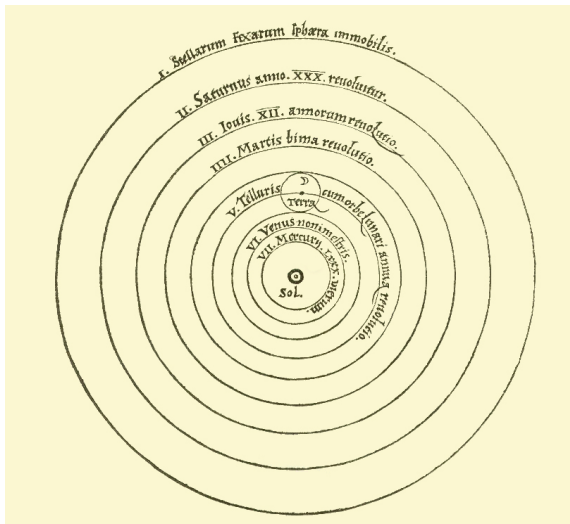
- Going beyond

Conclusion

Ptolemy v. Copernicus



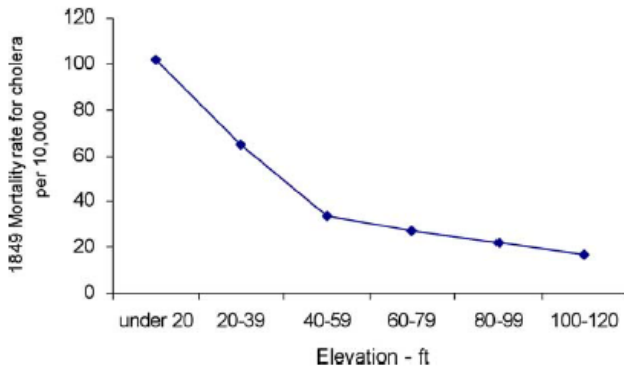
Ptolemy v. Copernicus



What causes cholera?



What causes cholera?



Outline

Conceptual models

Prediction

Model Validation

Model Evaluation

Goodness of fit

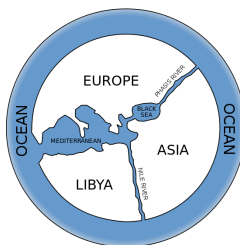
Capturing patterns

Going beyond

Conclusion

Model Validation

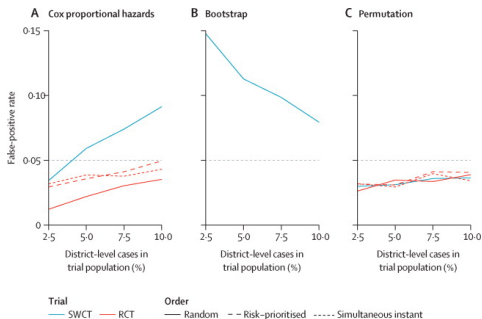
- ▶ Does your fitting algorithm match your *model world*?



- ▶ Coverage
- ▶ Precision
- ▶ Bias?
- ▶ Accuracy?

Coverage

- ▶ If you use your fitting algorithm on simulations from your model world, then you *know the right answer*!



- ▶ The right answer should be inside your 95% confidence interval 95% of the time
 - ▶ If more, your model is *too conservative*
 - ▶ If less, your model is *invalid*

Precision

- ▶ You should aim to make your confidence intervals as narrow as possible
 - ▶ Provide as much information as possible
- ▶ As data increases, your precision should increase
 - ▶ CIs should approach zero width

Bias?

- ▶ Nobody wants to be biased
- ▶ Good coverage and good precision, automatically ensures an *asymptotically* unbiased estimator
- ▶ Bias means that the mean expected prediction is the true value
 - ▶ Practical importance unclear
 - ▶ Scale dependent: an unbiased estimate of γ is automatically a biased estimate of D (but still can be asymptotically unbiased)

Accuracy?

- ▶ Nobody wants to be inaccurate
- ▶ Good coverage and good precision should guarantee good accuracy

Outline

Conceptual models

Prediction

Model Validation

Model Evaluation

- Goodness of fit

- Capturing patterns

- Going beyond

Conclusion

Model Evaluation

- ▶ Does your model match the *real world*?



Outline

Conceptual models

Prediction

Model Validation

Model Evaluation

- Goodness of fit

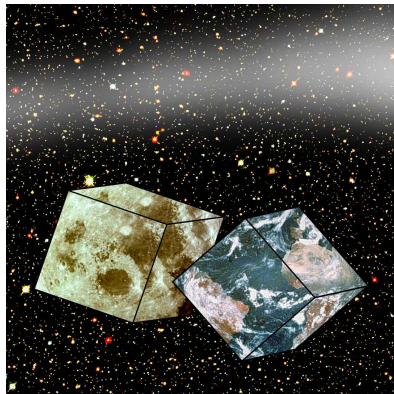
- Capturing patterns

- Going beyond

Conclusion

A disease-incidence model

Good for almost any disease

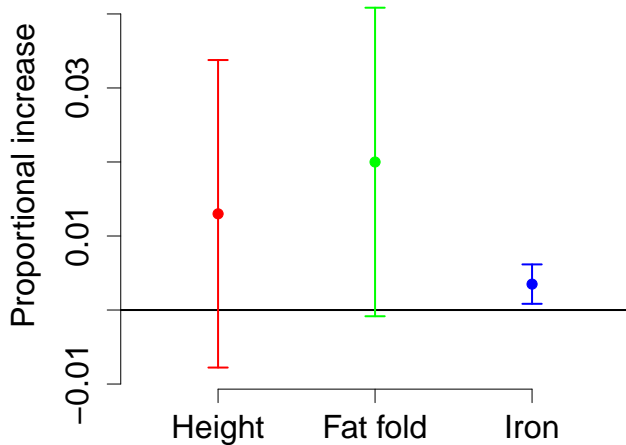


- ▶ The gods roll dice to pick a probability between 0.1% and 10%.
- ▶ Each person on the planet gets the disease the next year with this probability
- ▶ $P > 0.05$. My model is correct!

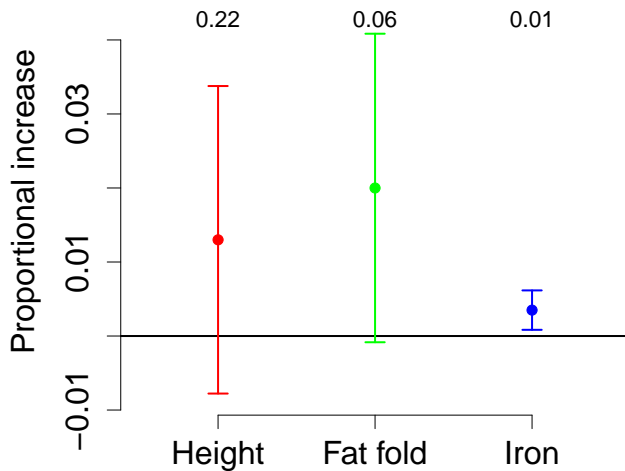
Your null model is false!

- ▶ So why do we use P values at all in biology?

Vitamin study



Vitamin study



Low P values



High P values



Goodness of fit test

- ▶ Your model is *not* reality (null hypothesis is false)
- ▶ Can we see the difference clearly?
 - ▶ If no, model may be good or bad. If good, we can't add any more complexity based on current data
 - ▶ If yes, model may be good or bad. We *may* be able to add more complexity based on current data
 - ▶ But we may not need to

Outline

Conceptual models

Prediction

Model Validation

Model Evaluation

- Goodness of fit

- Capturing patterns

- Going beyond

Conclusion

Capturing patterns

- ▶ You can ask:
 - ▶ Does your model do a reasonable job of capturing the data?
 - ▶ You can use a goodness of fit *statistic* for this, and not worry about the P value
 - ▶ Does your model capture patterns and relationships that you (or other experts) think are important?

Outline

Conceptual models

Prediction

Model Validation

Model Evaluation

- Goodness of fit

- Capturing patterns

- Going beyond

Conclusion

Out-of-sample validation

- ▶ Does your model make predictions *outside* the range on which you calibrated it?
 - ▶ Predicting gravitational shifts in star positions from measurements in Earth laboratories
 - ▶ Predicting cholera outbreaks in Bangladesh from a model calibrated to Haiti
 - ▶ Predicting influenza patterns in 2010 from a model calibrated from 2000–2009

Test sets

- ▶ What is **test set** spelled backwards?
- ▶ Hold some data out while fitting your model
- ▶ Or just *pretend* to do this as an evaluation method
 - ▶ In other words, test what would happen under various withholding scenarios

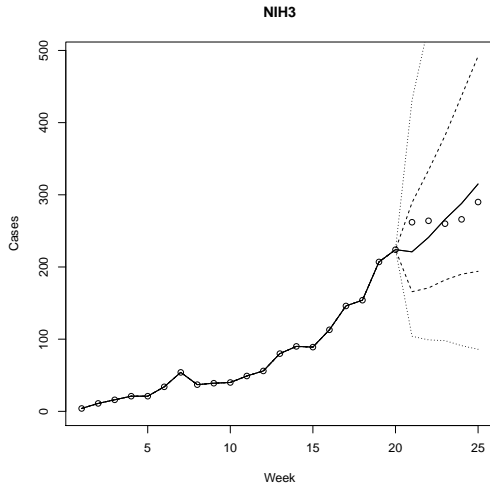
Other model worlds

- ▶ The model you're *fitting* is probably pretty simple
- ▶ But you can *simulate* very complicated models, indeed

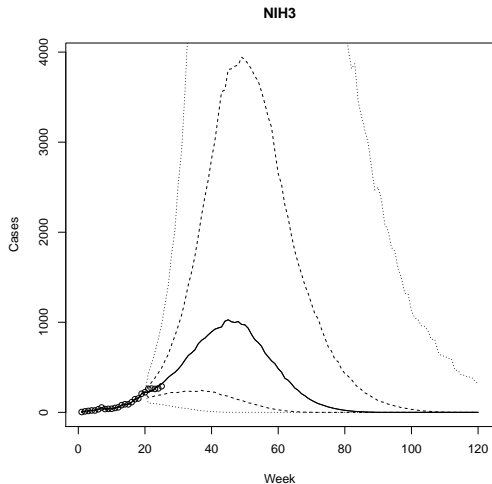


- ▶ How well can you do? Which details are important?

Other model worlds



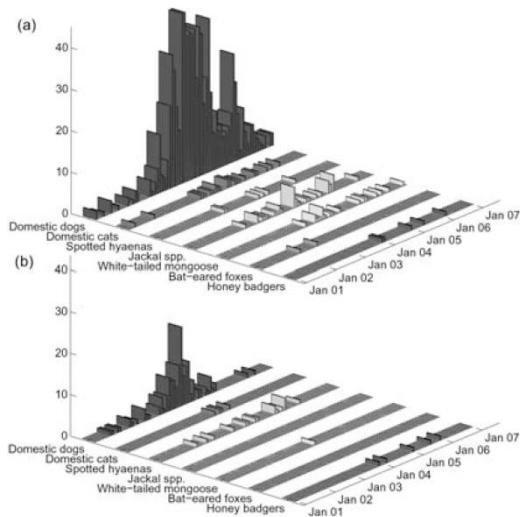
Other model worlds



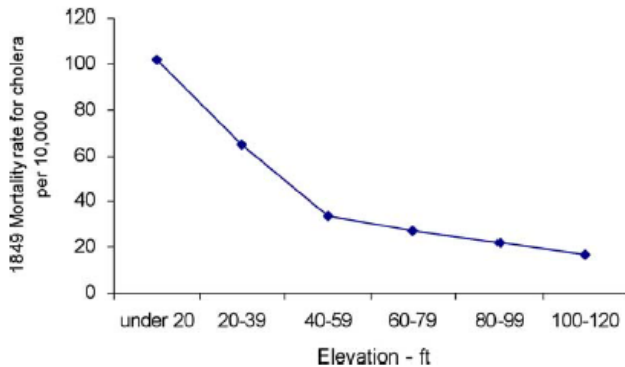
Generating hypotheses



Generating hypotheses



Testing hypotheses



Testing hypotheses



Testing hypotheses



Hard questions



Outline

Conceptual models

Prediction

Model Validation

Model Evaluation

- Goodness of fit

- Capturing patterns

- Going beyond

Conclusion

Dynamic models can help:

- ▶ Think clearly
- ▶ Understand outcomes
- ▶ Predict outcomes
- ▶ Find new mechanisms

Evaluation

- ▶ Validation (inside your model world)
- ▶ Inspection (compare patterns)
- ▶ Prediction (and other out-of-sample comparison)
- ▶ Generate and test hypotheses

Thank you

