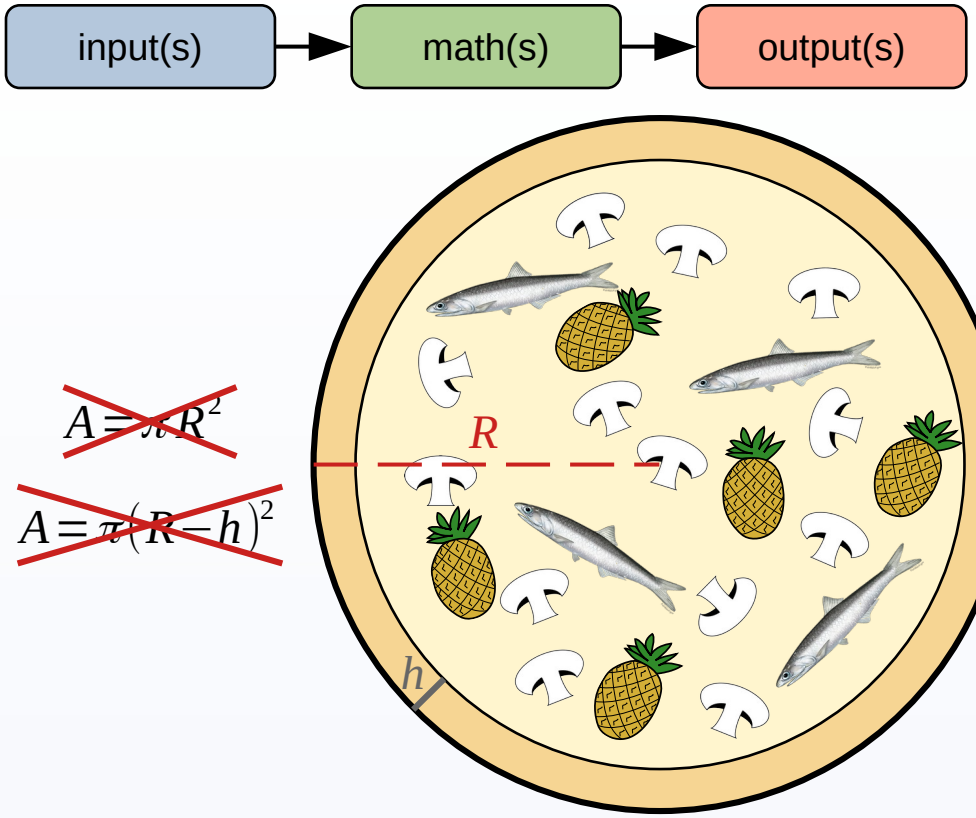
EGM101 – Skills Toolbox

Week 6, Part 6: Interpolation and Extrapolation

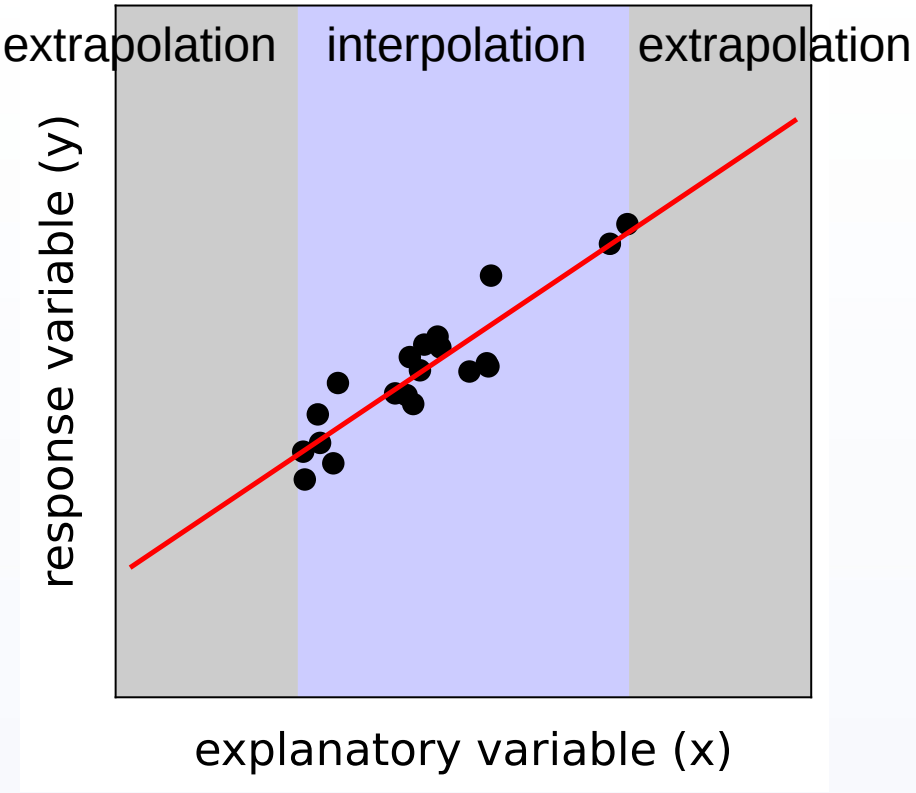
Recall: Mathematical models

- In science, seek to:
  - Understand (explain)
  - Make credible predictions
- One tool: mathematical models
- Example: surface area (A) of pizza
  - Simple!
  - But: what about crust?
  - But: crusts aren't even thickness, pizza isn't perfectly round, …
- Ultimately: "all models are wrong, but some are useful" (G. E. P. Box)



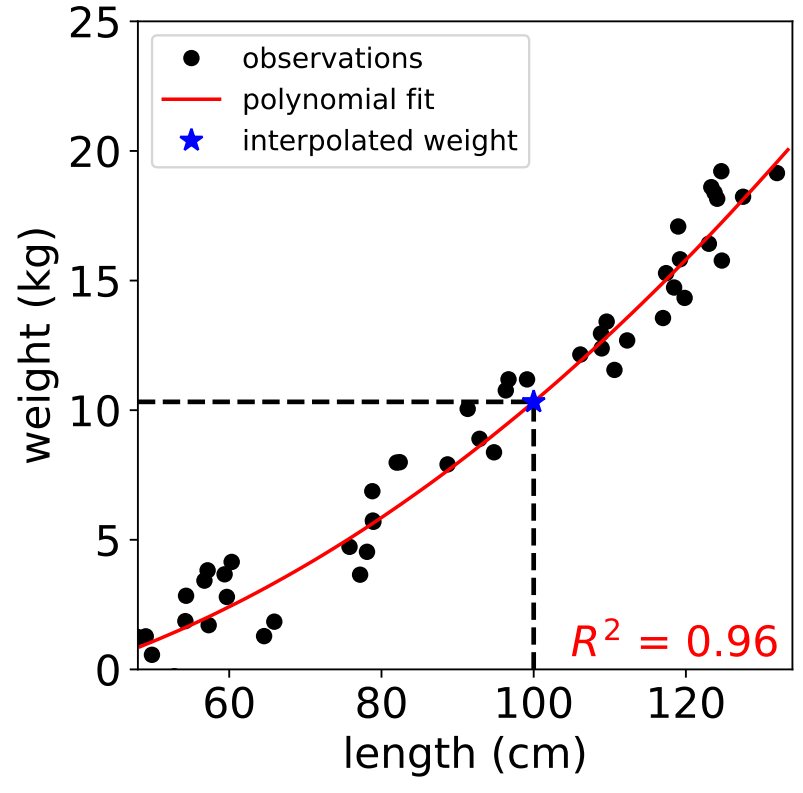
Definitions

- Interpolation: estimating unknown values of response variable within the range of observed (x) values
- Extrapolation: estimating unknown values of response variable outside of the range of observed (x) values



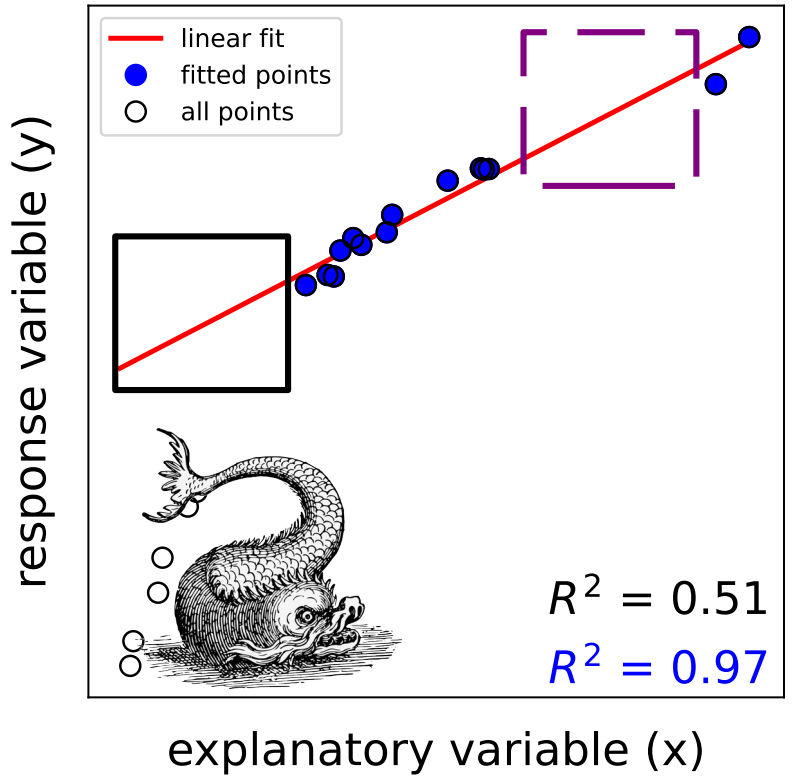
Interpolation Example: Salmometer

- Problem: how much does my fish weigh?
  - No scale
  - Catch-and-release
- In wild fish, weight tends to depend on length
- So, measure + weigh a number of fish, fit a relationship to the data
  - e.g., Salmometer



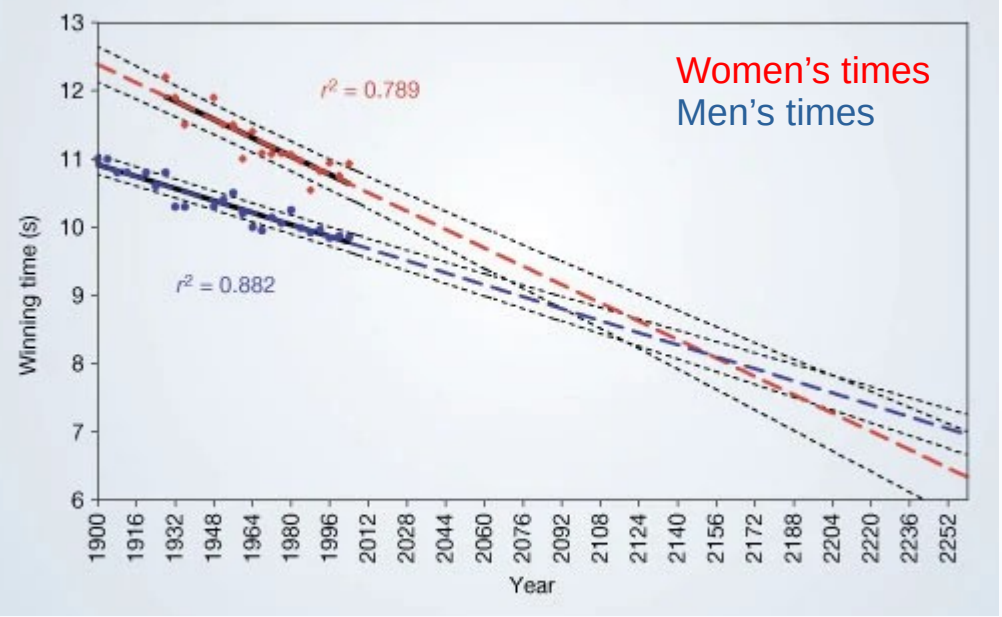
The dangers of extrapolation

- Interpolation: less likely that new observations completely contradict regression
- Can only be sure about “shape” of the relationship within the range of our observations
- Outside of this range, we don’t know what we don’t know.
  - Relationship could be non-linear
  - Could lead to ridiculous conclusions



A classic example

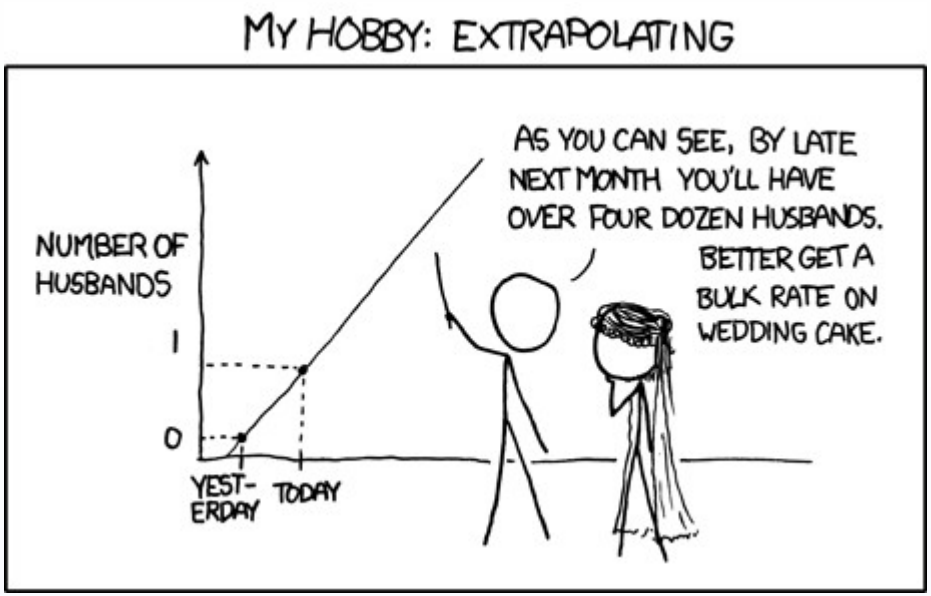
- Tatem et al., 2004:
  - Olympic winning times, 100 m (1900-2004)
  - Linear extrapolation shows that by ca. 2156, women's winning time will be faster than men's
  - Could be as early as 2064, as late as 2788 Olympic games
- Rice, 2004:
  - Model also predicts that in 2636, winning time < 0 s



Tatem et al., 2004

So when is it okay to extrapolate?

- It depends on what you are trying to model
  - Reasonable example: the sun will come out tomorrow
  - Unreasonable example: number of husbands over time.
- In general, it's good to have some kind of theoretical basis for your model first.



xkcd.com/605

Summary

- Part of using regression is to predict values:
  - Within the range of known values: interpolation
  - Outside the range of known values: extrapolation
- Interpolation is usually “safe”, as long as you have good data + a good model
- Extrapolation should be used sparingly, if at all.

Additional Resources

- Caswell, Chapter 9.4
- Weiss, Chapter 4.2
- Making Predictions with Regression Analysis [Jim Frost]
- Tatem et al. (2004) [Nature]
- Rice, K. (2004) [Nature]
- Salmometer: nfl.dfo-mpo.gc.ca/NL/AG/Salmometer