Model fitting

Peter Rowlett

Sheffield Hallam University

p.rowlett@shu.ac.uk

Models

- Look at the situations.
  - What would be the input variable? (think about what would go on the x-axis of a graph)
  - What would be the output variable? (think about what would go on the y-axis of a graph)
  - Think about any parameters that might be part of the model.
  - Think about the shape of graph you would expect if you plotted your output variable against your input variable.
  - Think about the class of mathematical function (e.g. linear, trigonometric, etc.) that would describe this relationship.

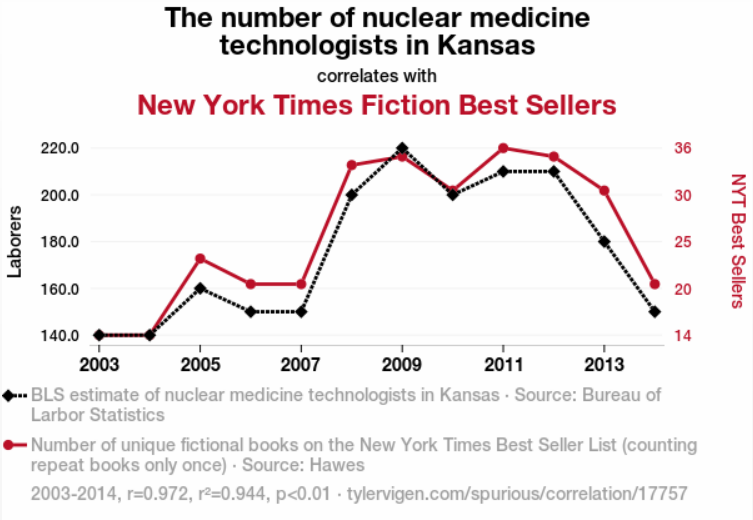
Models

- Look at the situations.
  - What would be the input variable? (think about what would go on the x-axis of a graph)
  - What would be the output variable? (think about what would go on the y-axis of a graph)
  - Think about any parameters that might be part of the model.
  - Think about the shape of graph you would expect if you plotted your output variable against your input variable.
  - Think about the class of mathematical function (e.g. linear, trigonometric, etc.) that would describe this relationship.
- Try to match the plot to the situation.

Models

- Look at the situations.
  - What would be the input variable? (think about what would go on the x-axis of a graph)
  - What would be the output variable? (think about what would go on the y-axis of a graph)
  - Think about any parameters that might be part of the model.
  - Think about the shape of graph you would expect if you plotted your output variable against your input variable.
  - Think about the class of mathematical function (e.g. linear, trigonometric, etc.) that would describe this relationship.
- Try to match the plot to the situation.
- Try to match the equation to the plot and situation.

A warning


The number of nuclear medicine technologists in Kansas
correlates with
New York Times Fiction Best Sellers
Laborers
NYT Best Sellers
220.0
200.0
180.0
160.0
140.0
36
30
25
20
14
2003
2005
2007
2009
2011
2013
BLS estimate of nuclear medicine technologists in Kansas · Source: Bureau of Larbor Statistics
Number of unique fictional books on the New York Times Best Seller List (counting repeat books only once) · Source: Hawes
2003-2014, r=0.972, r²=0.944, p<0.01 · tylervigen.com/spurious/correlation/17757

A warning

- This isn't a warning about employment rates for nuclear medical technicians.
- Just because two data sets move up and down together, doesn't mean there is a causal link.

What number to report?

- This article reports "There are many different estimates about the chances of such junk hitting someone, but most are in the one-in-10,000 range."
- This sounds unlikely, but is actually really alarming.
- In Census 2021, there were 556,500 people in Sheffield city. So we expect about 50 to be hit by space junk in . . . what timescale?

https://www.bbc.co.uk/news/articles/clyn9dgdwe3o

Space junk

- The BBC link to [a *Nature* article](https://www.nature.com/articles/s41550-022-01718-8).
- "In the USA, the Orbital Debris Mitigation Standard Practices (ODMSPs) apply to all launches and require that the risk of a casualty from a reentering rocket body is below a 1-in-10,000 threshold."
- So the 1-in-10,000 threshold is a target in the US, not an estimate of reality.

Space junk

- But also...
  "The 1-in-10,000 threshold for casualty risk is arbitrary and makes little sense in an era when new technologies and mission profiles enable controlled reentries. It also fails to address low-risk, high-consequence outcomes, such as a piece of a rocket stage crashing into a high-density city or a large passenger aircraft. In the latter case, even a small piece could cause hundreds of casualties."
- Anyway, the paper ends by saying "we conclude that current practices have on order a 10% chance of one or more casualties over a decade."

Which number to report?

- That's still quite alarming, but less so than 1-in-10,000.
- It's important to consider what you will report from your analysis, and how much that will make sense in the context you are modelling.

Does my model make sense?

- Formula for 'Blue Monday', reported over many years in different media outlets.

$$\frac{(W + (D - d)) \times T^Q}{M \times N_A}$$

where

- W is weather;
- d is debt;
- T is time since Christmas;
- Q is time since failing our New Year's resolutions;
- M is low motivational levels;
- N_A is the feeling of need to take action.

Does my model make sense?

- Just because your model produces a formula, doesn't mean it makes any sense.