WHAT ARE STATISTICAL MODELS, AND WHY DO I NEED THEM?

STEPHEN J. MILLER
WINCHESTER THURSTON SCHOOL
PITTSBURGH, PA

NCTM ANNUAL MEETING – SAN ANTONIO, TX
SESSION 208 – THURSDAY, 6 APRIL 2017

GOAL FOR TODAY

To give ***mathematics teachers*** a sense of how mathematical models describe the structure of a relationship, and how statistical models make them more thorough.

MATHEMATICAL MODELS

"A process that uses mathematics to represent, analyze, make predictions or otherwise provide insight into real-world phenomena."

Guidelines for Assessment and Instruction in Mathematical Modeling Education (GAIMME) 2016, p. 8

STATISTICAL MODELS

- Serve the same purpose
- Combine *mathematical models* with descriptions of *variability*

STATISTICAL MODELS

- Univariate (single-variable), bivariate (two-variable), and multi-variable statistical models take the form

data = *structure* + *variability*

STATISTICAL MODELS

- Include variability to provide a more thorough description of the behavior of the data under study.

UNIVARIATE STATISTICAL MODELS

- Univariate data values can be modeled as varying about a single numeric value, such as the mean (μ).
- The single value (μ) provides the *structure*.
- A statistical model incorporates *variability* in addition to the structure.

UNIVARIATE STATISTICAL MODELS

- A statistical model for each data value (y_i) might be:

$$y_i = \mu + \varepsilon_i$$

where ε_i represents a deviation from the mean for the i th data value.

- Deviations from the mean describe variability.

BIVARIATE STATISTICAL MODELS

- A statistical model for a bivariate ordered pair (x_i, y_i) might be:

$$y_i = f(x) + \varepsilon_i$$

where $f(x)$ is a function that describes the structure, and ε_i represents a deviation from the function $f(x)$.

HOW CAN WE EVALUATE STATISTICAL MODELS?

- Consider the ***usefulness*** and ***effectiveness*** in describing the data and making predictions
- We perform this analysis by considering how well the structural aspect models the data and by quantifying variability
  - Little variability, model is generally considered more useful than when there is greater variability

HOW CAN WE EVALUATE STATISTICAL MODELS?

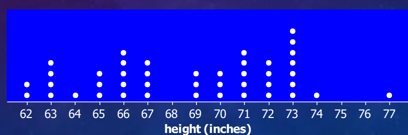
- With univariate data, we often use measures of variability such as ***standard deviation***, ***interquartile range***, or ***mean absolute deviation*** to measure how close a typical observation is from the center (mean or median)

HOW CAN WE EVALUATE STATISTICAL MODELS?

- With bivariate data, we have particular methods and measures to evaluate how well the structural component captures the overall pattern of the relationship.

MATHEMATICAL MODELS DESCRIBE STRUCTURE

- Consider a population of the heights of 43 statistics students.



MATHEMATICAL MODELS DESCRIBE STRUCTURE

- The mean (μ) is a commonly used measure of center

$$\mu = 68.7 \text{ inches}$$

- This is a mathematical model that ***summarizes*** the data.
- Describes a ***structural*** aspect of the population.

MATHEMATICAL MODELS DESCRIBE STRUCTURE

- This mathematical model is an ***incomplete*** description of the data.
- Why? Because the mean can describe the center of many different populations.

MATHEMATICAL MODELS DESCRIBE STRUCTURE

- Consider the following three populations:

MATHEMATICAL MODELS DESCRIBE STRUCTURE

- The mean alone only tells part of the story (the "structure").
- We need to somehow indicate that Population 1 has no variability in the data, and that the other two populations have some amount of variability about the mean.

MATHEMATICAL MODELS DESCRIBE STRUCTURE

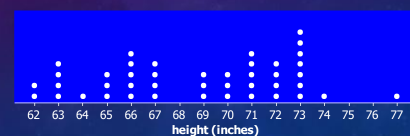
- Knowing only the structural component for the student height data ($\mu = 68.7$ inches) tells us nothing about the variability in heights.
- How do we summarize or measure how far values deviate, on average, from the mean?
  - The standard deviation

MATHEMATICAL MODELS DESCRIBE STRUCTURE

- For the heights data, the standard deviation is 3.84 inches.
- The standard deviation provides information beyond the structural component of the mean.

MATHEMATICAL MODELS DESCRIBE STRUCTURE

- We now know that, on average, the heights vary around the mean of 68.7 inches by about 3.84 inches



MATHEMATICAL MODELS DESCRIBE STRUCTURE

- Bivariate populations require functions of the form $y = f(x)$ to model structure, which is an overall ***trend*** or ***relationship*** between two quantitative variables.

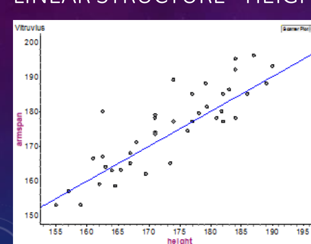
MATHEMATICAL MODELS DESCRIBE STRUCTURE

- Bivariate structural relationship include:
  - Linear
  - Quadratic
  - Exponential
  - Square root
  - Direct and inverse proportions
  - Inverse square relationship

MATHEMATICAL MODELS DESCRIBE STRUCTURE

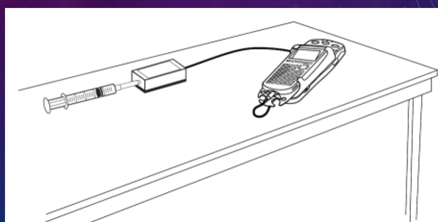
- The following slides show examples of ***student data collection activities*** that illustrate the structure of mathematical models.
- These data collection activities are relatively easy to do in class (less than one class period each).

LINEAR STRUCTURE– HEIGHT & ARMSPAN

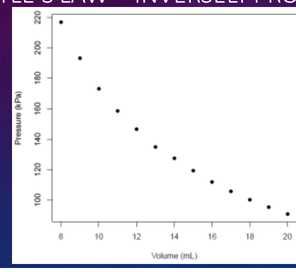


$height = armspan$

BOYLE'S LAW – INVERSELY PROPORTIONAL



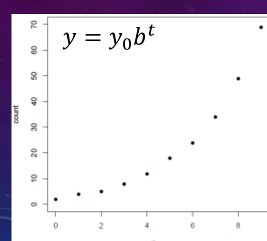
BOYLE'S LAW – INVERSELY PROPORTIONAL

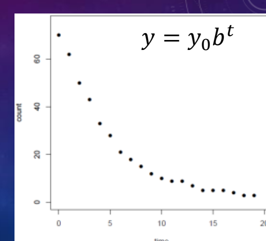


$$P = \frac{k}{V}$$

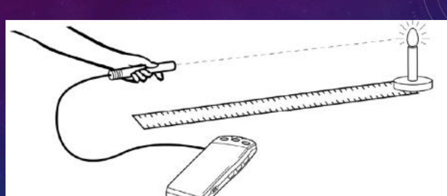
EXPONENTIAL GROWTH/DECAY WITH DICE

EXPONENTIAL GROWTH/DECAY WITH DICE

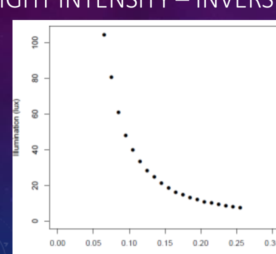




LIGHT INTENSITY – INVERSE SQUARE

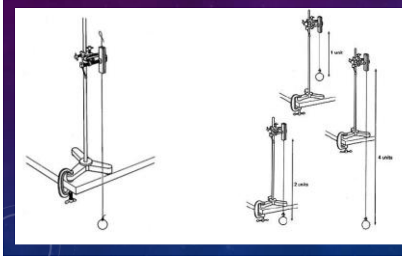


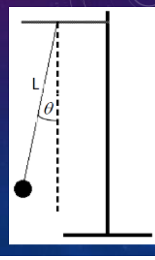
LIGHT INTENSITY – INVERSE SQUARE



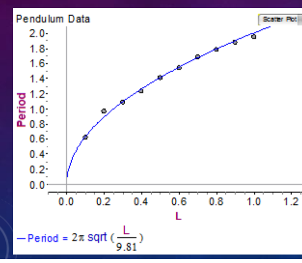
$$I = \frac{k}{r^2}$$

PERIOD OF A PENDULUM – SQUARE ROOT





PERIOD OF A PENDULUM – SQUARE ROOT



$$T = 2\pi\sqrt{\frac{L}{g}}$$

MATHEMATICAL MODELS DESCRIBE STRUCTURE

- These models don't take into account the variability that students are likely to see in their labs when they collect their own data.
- Variability arises from measurement error, from variations in the equipment, or variations in the settings in which measurements are taken.
- Theoretical models typically apply under certain conditions. Deviations from the conditions increase the variability.

GOOD MODELS GIVE USEFUL DESCRIPTIONS OF STRUCTURE

- Statistical models are evaluated by how well they describe data and whether they are useful.
- When used to describe physical phenomena, they are, *at best*, only approximations of reality.

GOOD MODELS GIVE USEFUL DESCRIPTIONS OF STRUCTURE

- Models are based on simplifying assumptions that have the benefit of producing a model that is relatively easy to understand, explain, and use.

GOOD MODELS GIVE USEFUL DESCRIPTIONS OF STRUCTURE

- Let's consider the structural model for the period of a pendulum in more detail.

$$T = 2\pi\sqrt{\frac{L}{g}}$$

GOOD MODELS GIVE USEFUL DESCRIPTIONS OF STRUCTURE

- This is a theoretical model of the structure of the relationship between period and length.
- This model is for a point pendulum supported by a massless, inextensible cord.
- This model applies when the angular displacement of the pendulum is no more than 10 to 15 degrees.

GOOD MODELS GIVE USEFUL DESCRIPTIONS OF STRUCTURE

- Students collect data to investigate the relationship using real (*i.e.*, non-point) pendulum bobs and cords with both mass and the potential to stretch.
- Result: a scattering of points about the curve for the mathematical (structural) model.

GOOD MODELS GIVE USEFUL DESCRIPTIONS OF STRUCTURE

- In addition to the unrealistic assumptions of the theoretical model, other deviations from theoretical results could be due to:
  - Buoyancy of pendulum bob in air.
  - Air resistance as the bob swings.
  - Inaccuracies in measurements of length and period.

GOOD MODELS GIVE USEFUL DESCRIPTIONS OF STRUCTURE

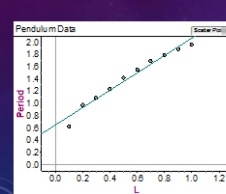
- Statistical models (*also approximations of reality!*) extend mathematical models and allow for incorporation of variability around the structure.
- The statistical model is:

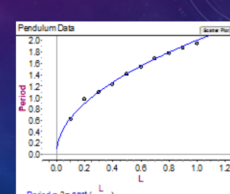
$$T = 2\pi\sqrt{\frac{L}{g}} + \varepsilon$$

GOOD MODELS GIVE USEFUL DESCRIPTIONS OF STRUCTURE

- Given that a statistical model is also an approximation of reality, how should we evaluate whether or not it is a good model?
  - How well does the structural part of the model describe the data?
  - How useful is the model?

GOOD MODELS GIVE USEFUL DESCRIPTIONS OF STRUCTURE





GOOD MODELS GIVE USEFUL DESCRIPTIONS OF STRUCTURE

- The line captures the increasing trend in the data (longer pendulums have longer periods).
- However, there are ***systematic deviations*** from the line.
- These systematic departures from the model indicate that the line doesn't accurately capture the structure of the data.

GOOD MODELS GIVE USEFUL DESCRIPTIONS OF STRUCTURE

- The variability in the statistical model is not intended to describe the structural component of the model.
- The variability describes the non-systematic deviations from the structure.

GOOD MODELS GIVE USEFUL DESCRIPTIONS OF STRUCTURE

- A good model will ideally have small variability with respect to the structure.
- Small variability translates into an improved ability to make predictions about future observations.

GOOD MODELS GIVE USEFUL DESCRIPTIONS OF STRUCTURE

"The most that can be expected from any model is that it can supply a useful approximation to reality:

All models are wrong;
some models are useful."

Box, Hunter, and Hunter
Statistics for Experimenters

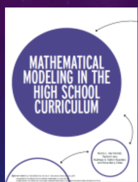
GOOD MODELS GIVE USEFUL DESCRIPTIONS OF STRUCTURE

- Statistical and mathematical models are fundamentally simplifications of reality.
- The structural component is often based on assumptions that reduce the complexity of the relationship.

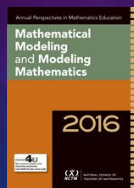
GOOD MODELS GIVE USEFUL DESCRIPTIONS OF STRUCTURE

- Models become useful when we can use the structural part to accurately and precisely describe the data, taking variability into account.
- This can be accomplished more successfully for data with less variability than for data with more variability.

RECOMMENDED RESOURCES



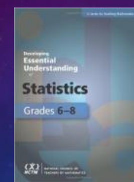
Mathematics Teacher
December 2016/January 2017

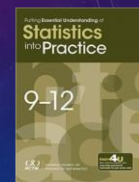


Annual Perspectives in Mathematics Education

Guidelines for Assessment and Instruction in Mathematical Modeling Education

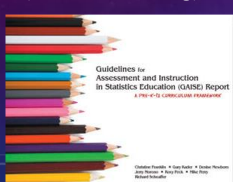
RECOMMENDED RESOURCES

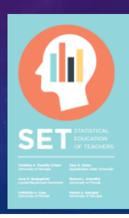




RECOMMENDED RESOURCES

- From the American Statistical Association (www.amstat.org):





QUESTIONS AND CONTACT

- Questions?
- Feel free to contact me:
  - Stephen Miller
  - sm1016@gmail.com
  - Pittsburgh, PA