Knowledge for Teaching Informal Line of Best Fit

2014 NCTM Research Conference

New Orleans, LA

April 8, 2014

Research Team

PI: *Stephanie Casey, Eastern Michigan University*

Statistical knowledge for teaching study:

- Tim Burgess, Massey University (New Zealand)
- Julie Clark, Hollins University
- Jennifer Kaplan, University of Georgia
- *Adam Molnar, University of Georgia*

Research Team

Student conceptions study:

- David Wilson, SUNY Buffalo

Teacher knowledge study:

- *Nicholas Wasserman, Teachers College, Columbia University*

Discussant:

- *J. Michael Shaughnessy, Portland State University*

Informal Line of Best Fit (ILOBF)

- Statistical association identified as one of eight big ideas in statistics (Garfield & Ben-Zvi, 2004)
- ILOBF: first time students study statistical association
- Included in statistics curriculum worldwide, including Australia, Brazil, England, & U.S.A.

ILOBF in the CCSS-M

8.SP.A.2: Know that straight lines are widely used to model relationships between two quantitative variables. For scatter plots that suggest a linear association, informally fit a straight line, and informally assess the model fit by judging the closeness of the data points to the line.

8.SP.A.3: Use the equation of a linear model to solve problems in the context of bivariate measurement data, interpreting the slope and intercept.

Thinking about the informal line of best fit

- Using the piece of spaghetti, determine the line of best fit for the data shown in each scatterplot.
- Be cognizant of your thoughts as you decide where to place the line on each graph.

Thinking about the informal line of best fit

- What things did you consider when you were deciding where to place it?
- Why did you choose to put the line there?
- What is your definition of the line of best fit?

What do teachers need to know in order to help students do this well?

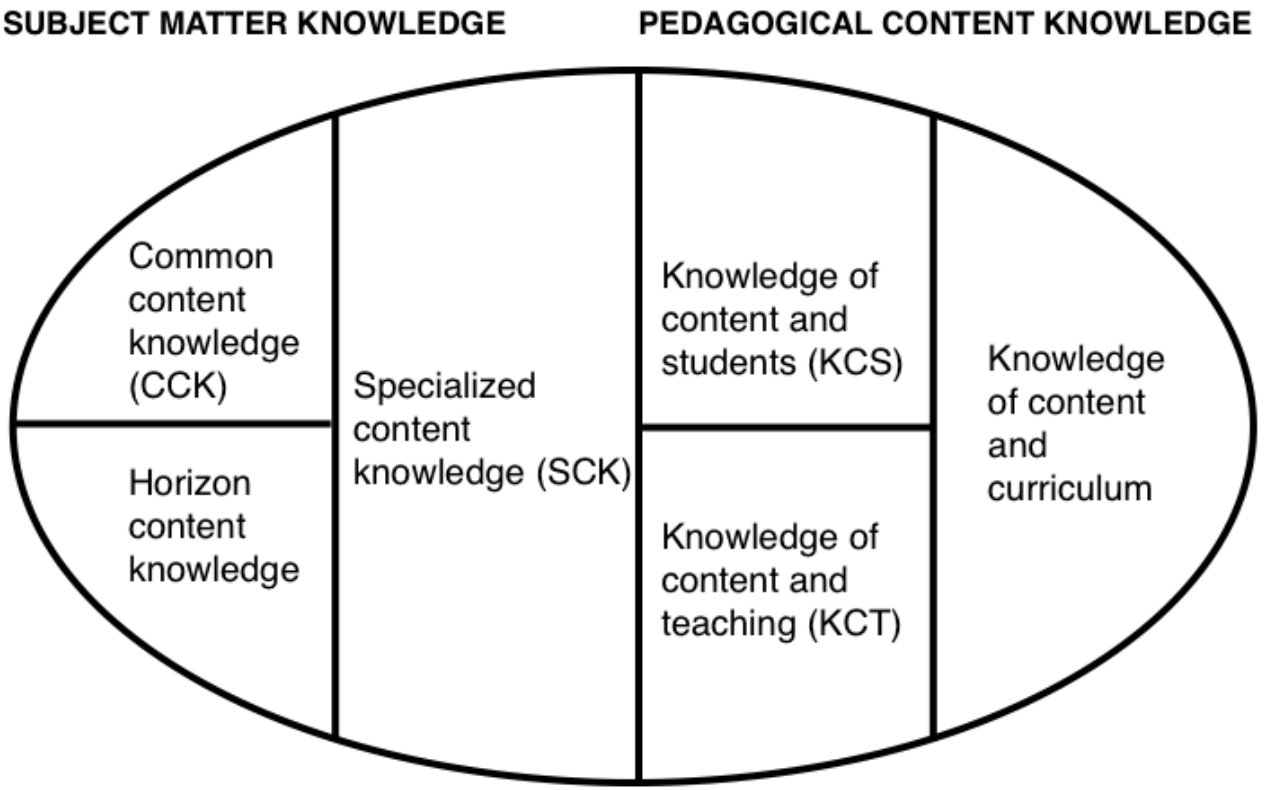
Why study knowledge for teaching ILOBF?

- Prominence in curriculum worldwide
- Opportunities it provides for students to create their own methods
- Topic for which differences between mathematical and statistical reasoning are noticeable and can be studied
- Field generally lacks descriptions of teacher knowledge for teaching statistics

Research Question

What is the knowledge
needed by teachers
to teach
informal line of best fit?

Mathematical Knowledge for Teaching Framework (Ball, Thames, & Phelps, 2008)



Practice-based approach

Studied a teacher involved in the work of teaching ILOBF with an analytical focus on the demands upon teachers' knowledge involved in the practice of teaching

Design of the Study

- Generation of data at both *primary* and *secondary* levels
- Primary: case study of a secondary mathematics teacher as she taught ILOBF over five class sessions
- Secondary: analysts' conjectures regarding the needed knowledge using the MKT framework

Results categories

- *Model*
  - Choosing appropriately
  - Differentiating from functions
- *Purpose & Placement*
  - Purpose
  - Placement and Criteria
- *Scatterplots*
  - Graphing
  - Reading
- *Predicting*
  - From different representations
  - Variability in

Modeling: Linear Functions comparison

- Function: each input has exactly one output; Data may not
- Function: monotonically increasing or decreasing; Data may not be
- Function: graph goes through every point in table; Model does not

Modeling: Slope

	Mathematical Function	Statistical Model
Number type	Fraction	Decimal
Emphasis on	Rise over run	Average increase in response variable for a one-unit increase in predictor variable
Interpretation	Deterministic: will always happen every time	Average or predicted
Zero slope	Special case; minimal exposure	Important case: no association between variables

$$y = \frac{3}{5}x - 2$$

Bounce height = 0.7 Drop Height -3.4

Modeling: y-intercept

- Mathematical function: usually of interest; often starting point when graphing
- Statistical model: usually not of interest; when fitting line of best fit, it is not the starting point

Bounce height = 0.7 Drop Height -3.4

Modeling: PCK

- Students' knowledge of mathematical linear functions can interfere with their understanding of data and models
- Choose and sequence examples that help students understand differences between statistical models and mathematical functions

Modeling: PCK

- Organization of student learning outcomes for linear functions and linear models within the curriculum
- Types of models covered in school curriculum

Case-oriented vs. Aggregate Views

(Bakker, 2004)

- Novel learners take a case-oriented approach: perceive data as a series of individual cases, focus on a few points
- Students need help developing aggregate reasoning: analyze the data as a whole entity with characteristics that are not visible in any of the individual cases
- Particularly relevant to line of best fit

Purpose of Line of Best Fit

- ILOBF as an aggregate model for data
- Signal amid noise conception: Finding a line of best fit is a process that models the pattern (or signal) that can be found in the data once we have accounted for the variability (or noise) (Konold & Pollatsek, 2002)

Purpose of Line of best fit

- Explain sources of variability/noise in the data
- Student difficulties with aggregate reasoning
- Connect use of linear models as an aggregate model for bivariate data to other uses of aggregate reasoning in statistics like the mean of a univariate data set

Placement of ILOBF

- Teacher needs to provide statistically accurate, responsive, real-time feedback to students as they create their own lines
- Ability to estimate location of linear trend with precision
- Knowledge of the procedures used to place linear models

Purpose & Placement of ILOBF

- Anticipation of student difficulties with multiple solutions when placing ILOBF
- Knowledge of student conceptions about purpose & placement of lines of fit

MISSING FROM LITERATURE

Student Conceptions of ILOBF

- What do students conceive the line of best fit to *be*?
- What are students' natural inclinations regarding *criteria* the line of best fit is trying to meet and what *method*s do they use to meet them?
- What are *sources* of students' conceptions?

Student Conceptions of ILOBF

- How accurately do they place ILOBF?
- How do they interpret a placed ILOBF in context?

Student study methodology

- Task-based interviews with 33 eighth grade US students
- 6 tasks and related questions
- 5 tasks placing a line
- 6th task presents two students' lines of best fit & asks which line is better

Student meanings for LOBF

- Where you expect the relationship between the variables to be
- Shows what the data looks like
- Average
- Something you use to get close predictions

Student criteria

- Selected criteria presented in order from most common to least common
- Video clips
- Many students wanted to bend wire to connect points

Through as many points as possible

S2 3:07-3:18

- Connection with prediction definition

Through origin + most points

S31 1:50-2:31

Alright, so when I have it here there's 2 dots on the wire, so...if I move it slightly I get 3 or 4 actually. 1,2,3,4, OK sorry*. I'll let you hold it in place, I'll just tape it down. Alright, it looked to me like you were holding this corner down here (points to origin) in place and just moving the one end. Is that right?* Yes *Why did you hold this end (points to origin) in place?*Um, I wanted to make it a [unintelligible] variable where it starts at (0,0) 'cause if you don't drop it then it's not going to go up.

Through origin + most points

2 reasons cited for forcing line to go through the origin:

- Context
- Where all lines start

Equal number

S12 7:28-7:51 (accurate)

Equal number

S15 7:04-7:48+8:22-8:25 (inaccurate)

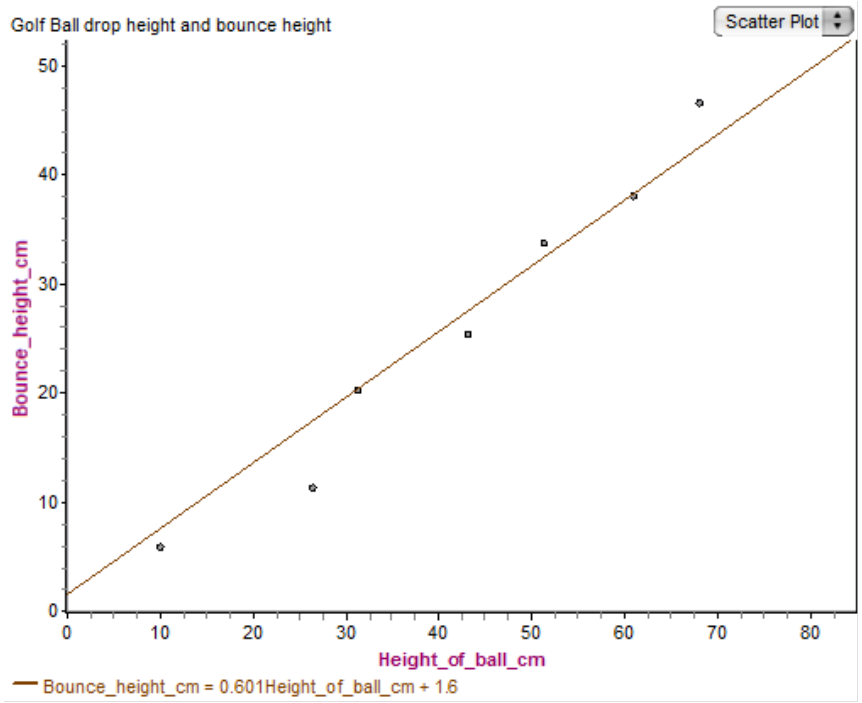
- Connection with average (median) and expect relationship definitions

Closest to all points

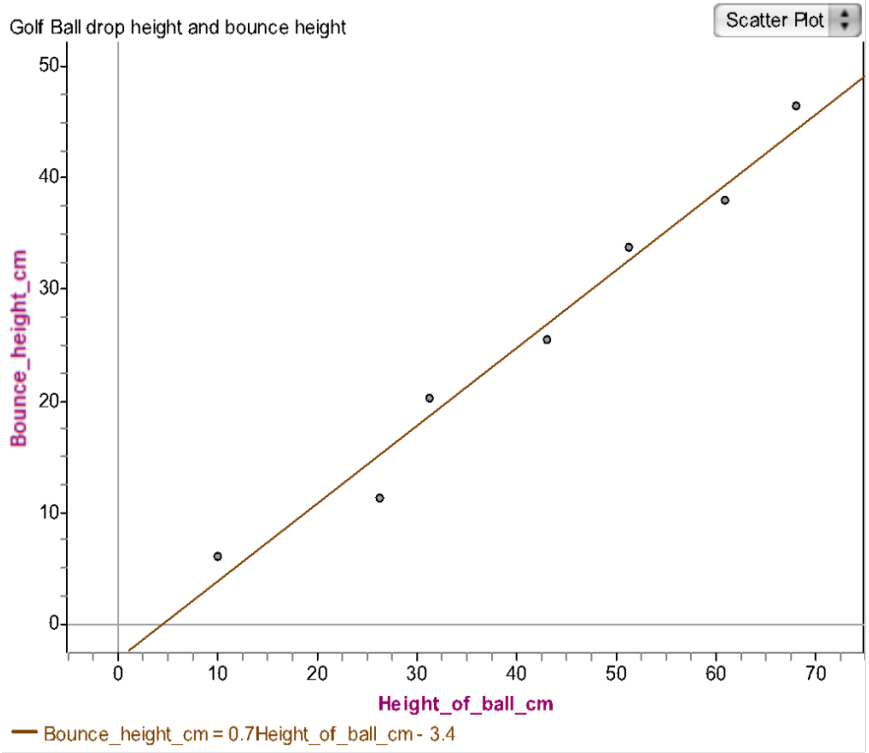
S3 7:46-8:35

Task 6: Which LOBF best?

Line A



Line B



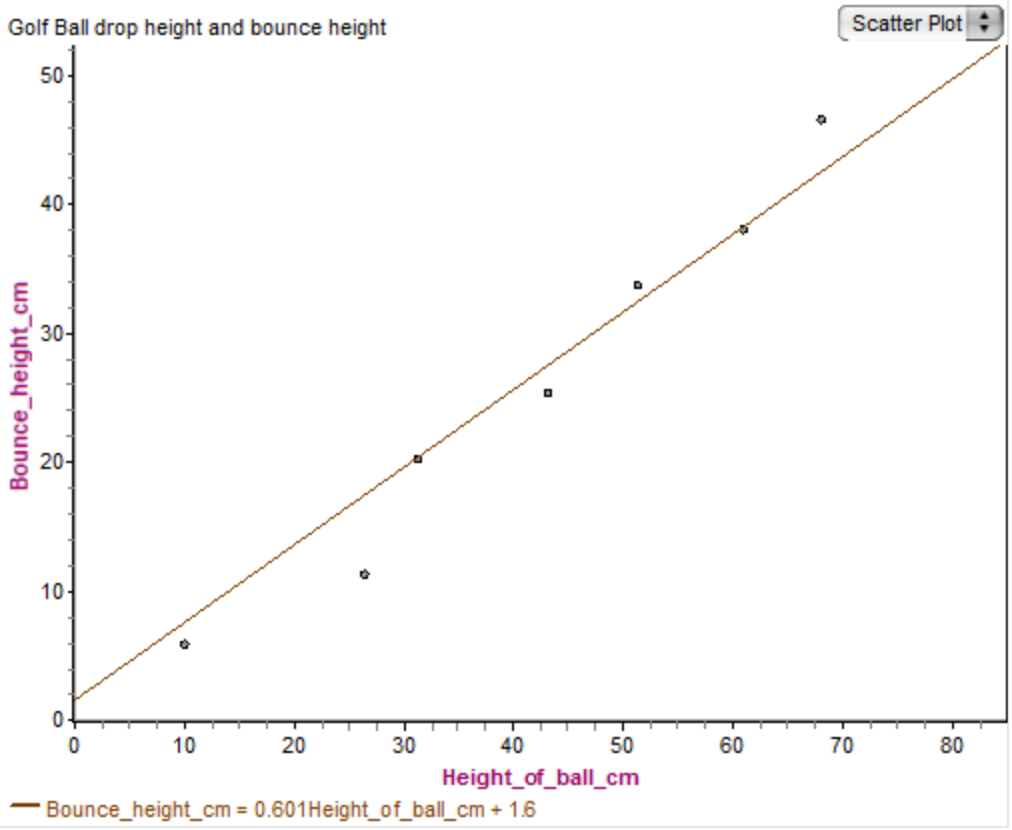
Line A better

S9 17:40-18:24

Line B better

S9 17:40-18:24

1/3 of students preferred Line A to Line B (LSRL)



Aggregate view

- going through the most dots
- starting at the origin then going through the most dots
- equal number of dots on both sides of the lines (could be always vertical or always horizontal)
- as close to all of the points as possible (CCSS-M)

Case-oriented view

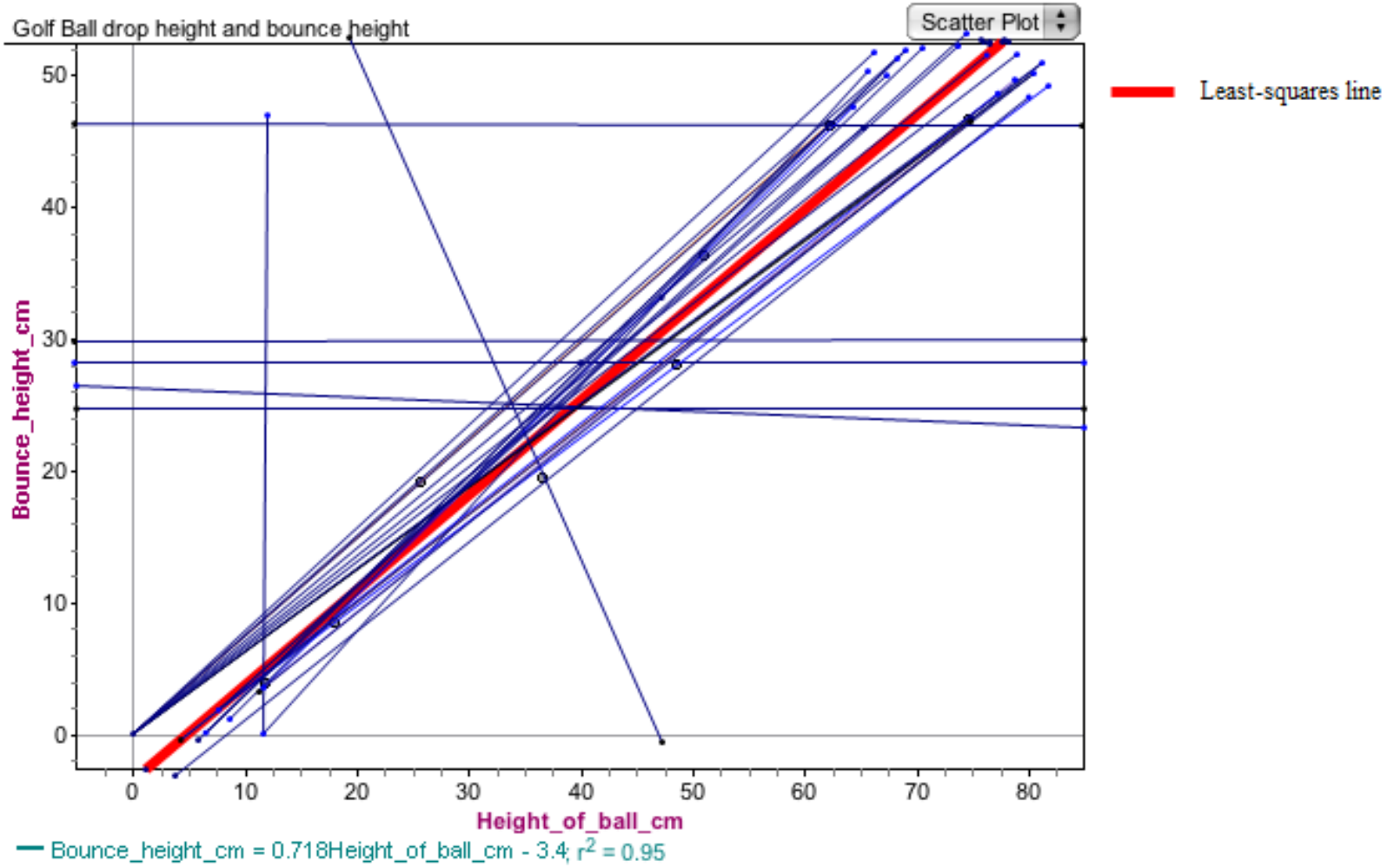
- Halfway between the highest and lowest points
- Through first and last point

Case-oriented view

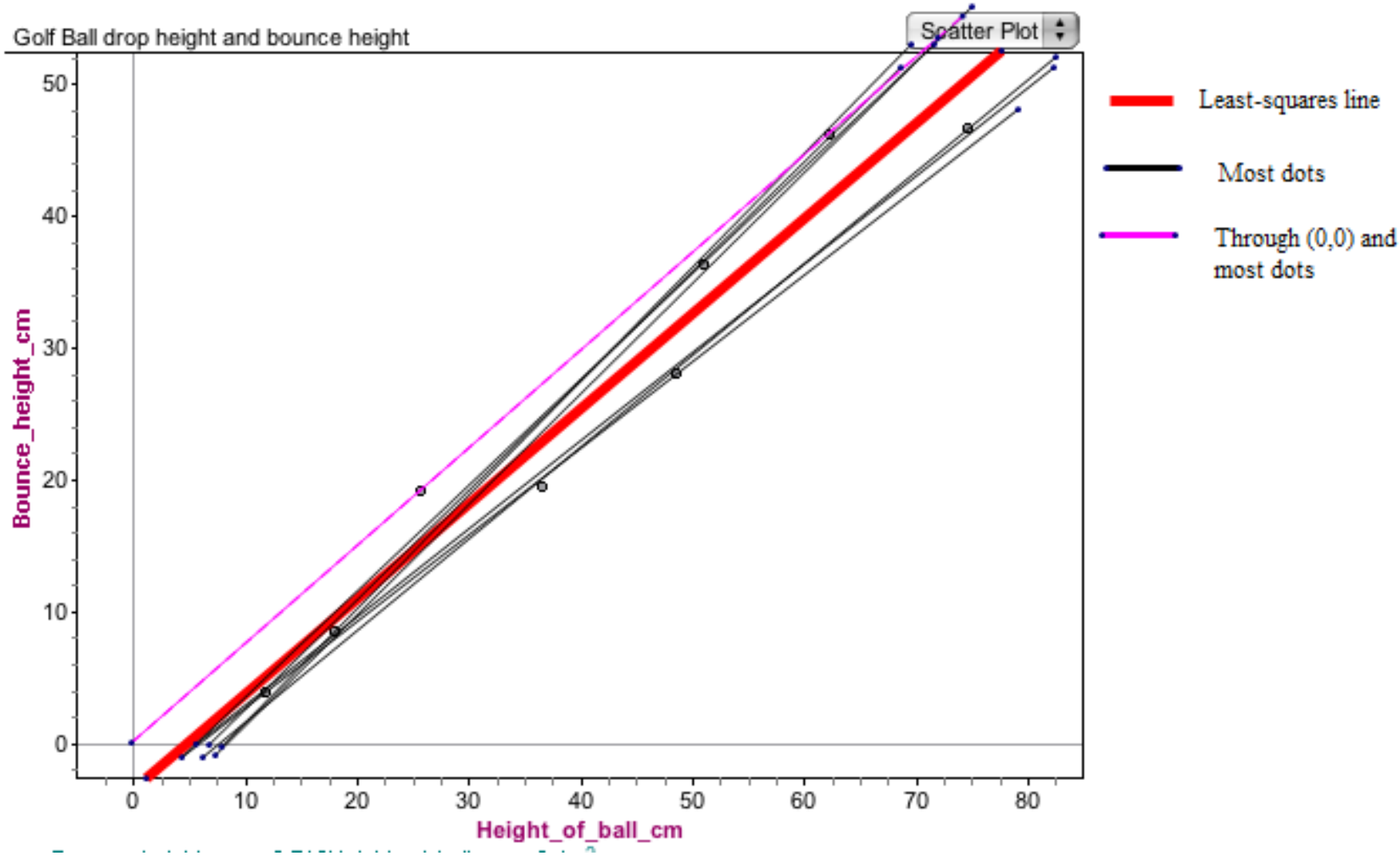
- Through points with the same y-coordinate

S16 7:38-8:25

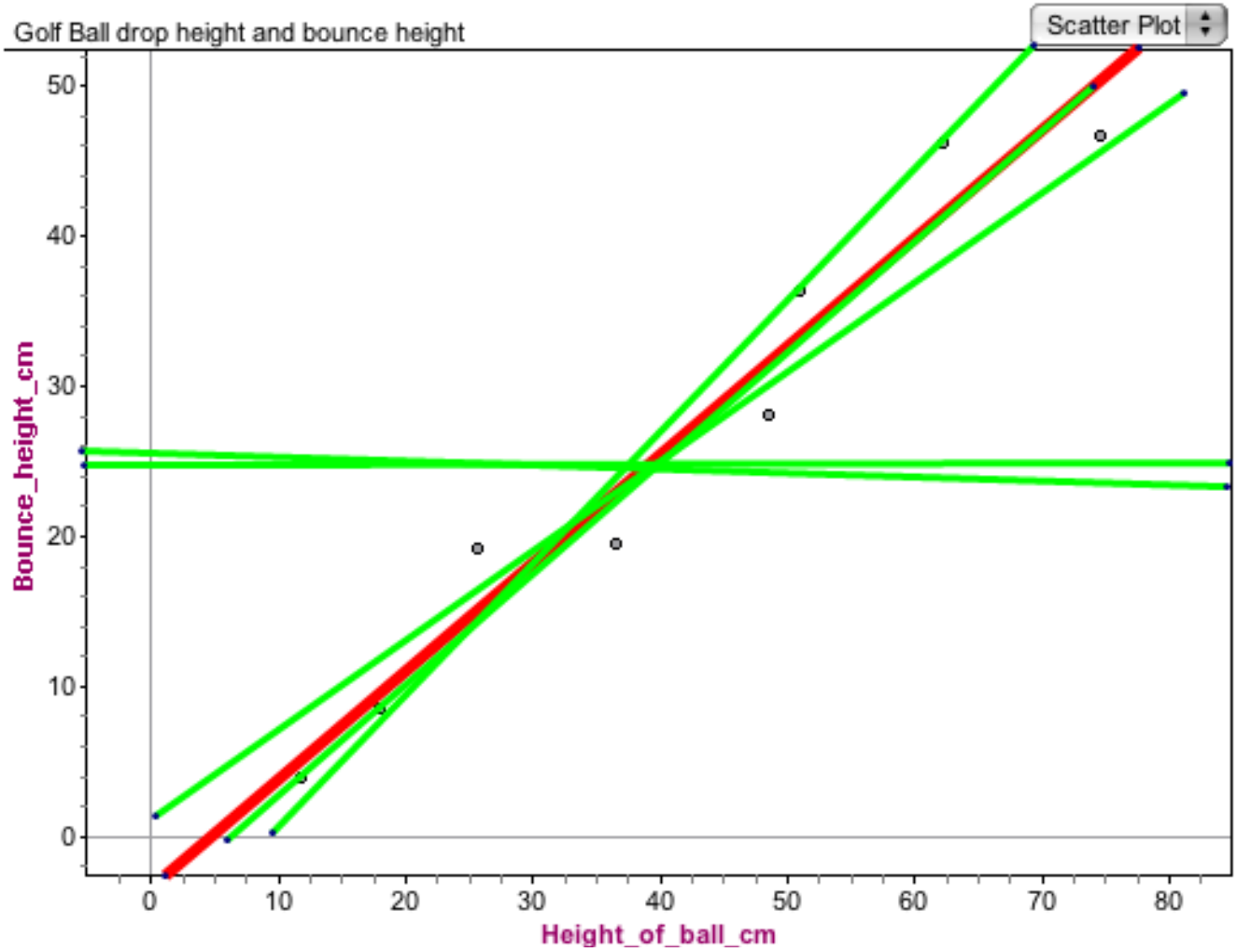
Student line placement: Task 2



Most dots lines only



Equal number lines only



Implications for the field

- Findings provide teachers and teacher educators with descriptions of student conceptions about purpose and placement of lines of fit to develop their KCS for teaching LOBF
  - E.g. Likely that a sizeable number of students will think going through points is more important than closest to all points
- Findings inform school curriculum writers

  E.g. Choices of scatterplots for students to fit lines to

Teacher knowledge study

- Added a teacher level study to parallel the student level study
- Greater inclusion of informal regression in school curricula presents the need for research on teacher knowledge about informal regression
- Focus on subject matter knowledge

Teacher study methodology

- Task-based interviews with 19 teachers
- Teachers enrolled in teacher education courses at 3 large US universities
- 11 preservice, 8 inservice
- Interview protocol like student study with addition of questions regarding interpreting slope and causation

Teacher meanings for LOBF

Meaning	Description (The line of best fit exemplifies…)	Example excerpt	Number of participants
Model	The idealized/theoretical relationship between the variables.	"it is just a line that, um, as closely approximates the actual true relationship as possible." [T7]	5
Signal	The meaningful and general signal for data that eliminates the inherent noise.	"the line of best fit is…something that's uncomplicated…something simple and easy to visualize, [trying] to find a relationship…" [T5]	2
Typical	The bivariate equivalent for determining middle or typical value; sometimes talked about it as average, other times as median.	"The line that averages all the – the average of the points." [T6]	3
Representer	The best representation of all of the data, where the line accounts for the data at hand rather than a more general relationship.	"…a line that represents the set as a whole as opposed to any particular data point." [T1]	7
Predictor	The line that enables predictions for values not in the data set.	"so we can predict where it might be in the future... gives you an idea where the answer is going." [T17]	2

Teacher criteria

- 4 dominant
- Video clips

Equal number

- EMU T9 8:30-56, 9:19-10:05

Pairs

- EMU T2 13:45-14:42

Closest to all points

- SUNY T2 4:15-4:35

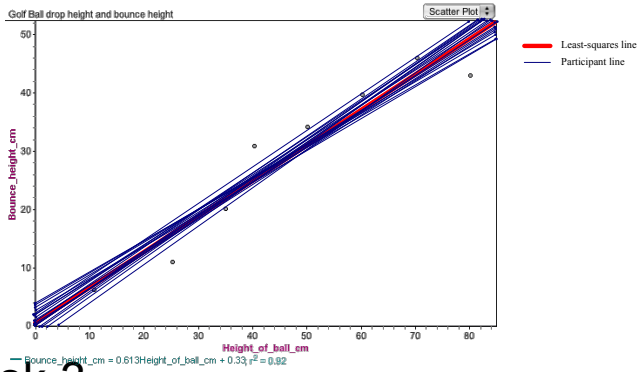
Equal sum deviations

- EMU T5 7:35-7:46

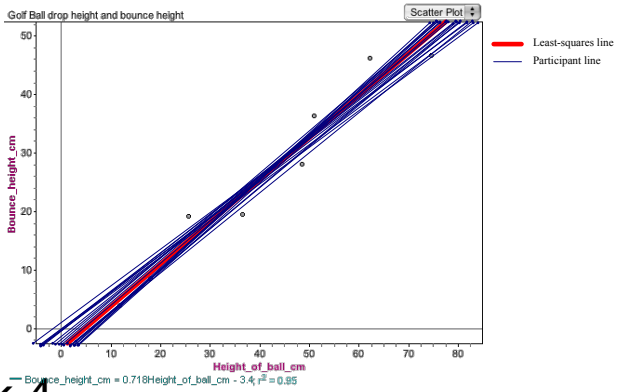
Placement of ILOBF: Tasks 1-4

- Despite differing criteria, notably similar lines

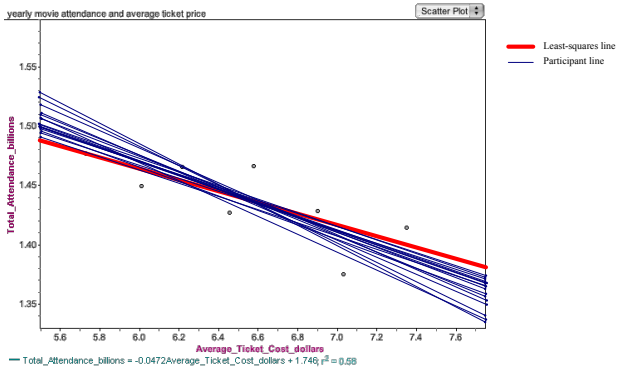
Task 1



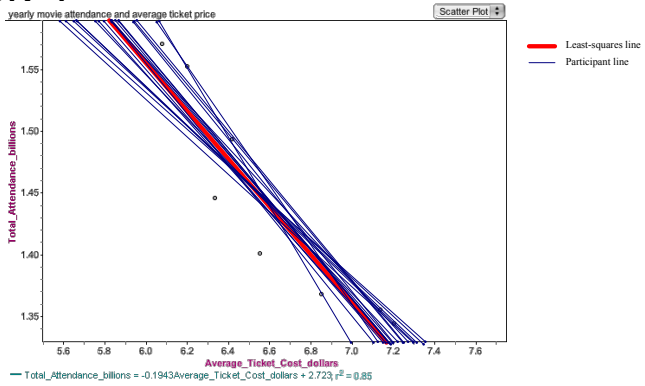
Task 2



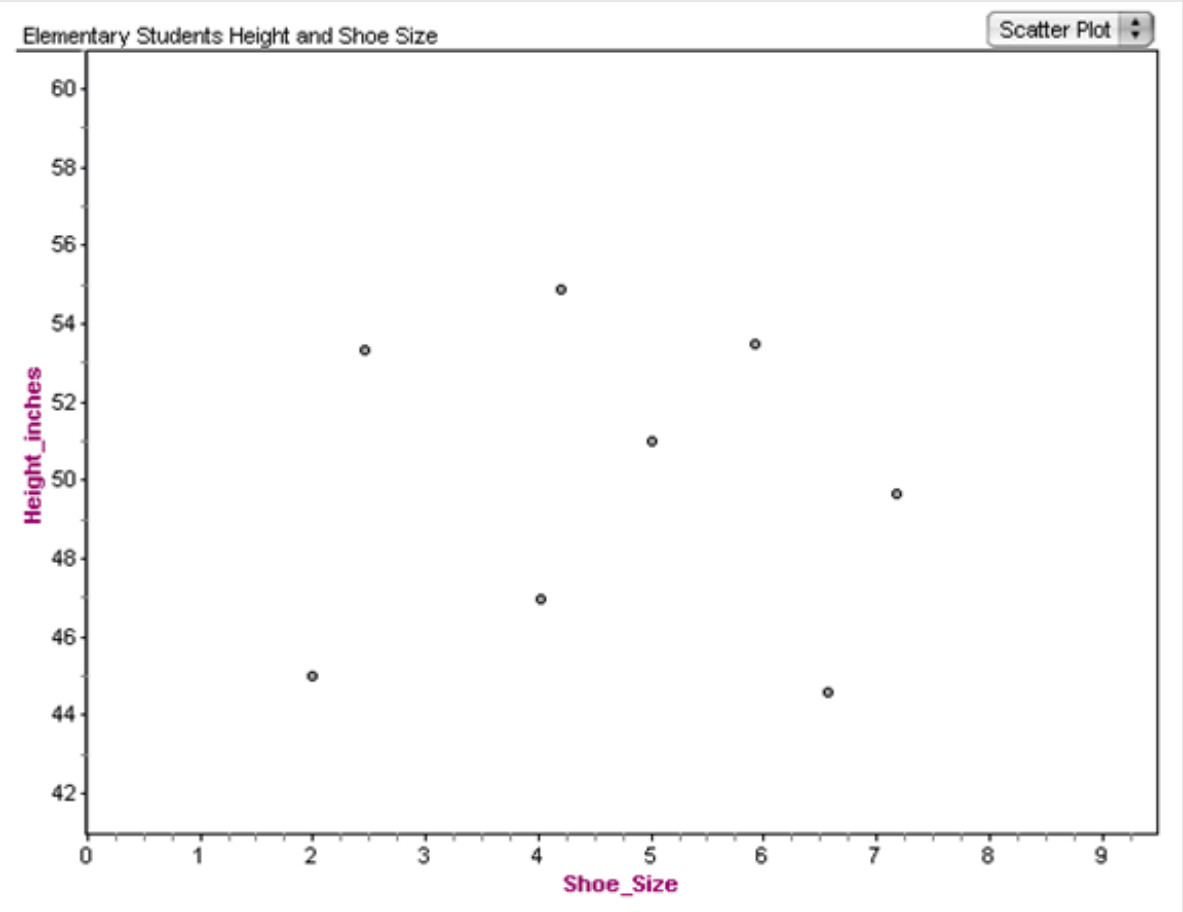
Task 3



Task 4

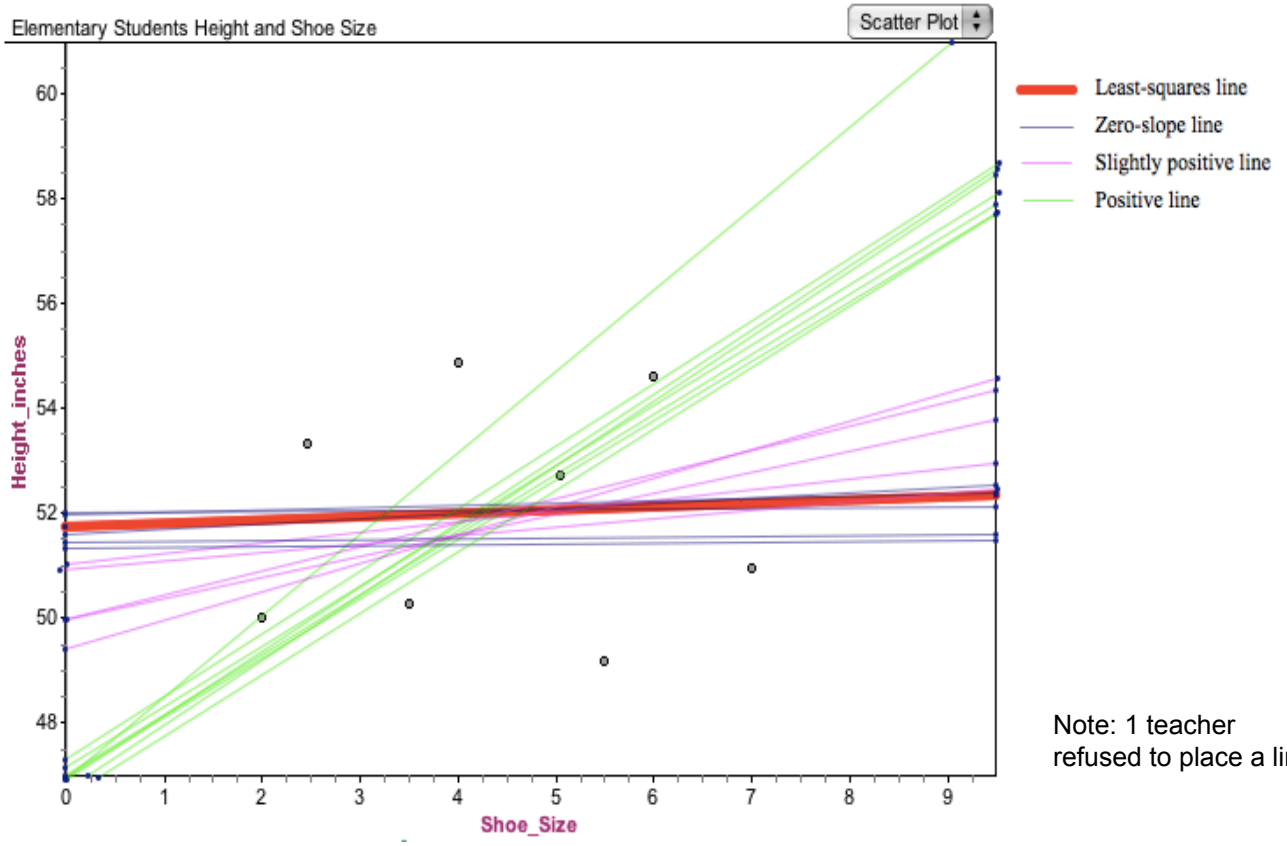


Task 5: No association



Task 5: No Association

- More variation: 3 main categories



Note: 1 teacher refused to place a line

No association task

- Further findings
  - Dominant criteria: P and SP groups mostly altered their criteria for this task; Z did not
  - Context: Those mentioning context as important to ILOBF was significantly associated with P and SP groups ($p=0.045$)
  - Subset: Those who ever relied on a subset of data for ILOBF was also significantly associated with P and SP groups ($p=0.033$)

No association task

- Discussion
  - Fitting line to non-associated data only makes sense through statistical inference (not associated)
    - Struggles potentially indicative of lack of HCK
    - Teachers' criteria & conceptions need to still make sense with non-associated data
  - Math classes often de-emphasize zero & undefined slopes with linear functions (e.g., y=2 or x=-3). But implications about the association between x & y is important, particularly in statistics.

Further findings

- Mathematics and Statistics
  - Confused more deterministic approaches and more flexible approaches in math/stats
    - Considered context as essential when it was not (least squares LOBF is context-free procedure, deterministic)
    - Had more deterministic interpretations of slope (when less rigid interpretation is necessary)
- Association vs. Causation
  - Most teachers incorrectly assumed causation, but teachers who had taught LOBF did not.

Implications for the field

- Implications from no association task
  - Teachers need to be more knowledgeable of surrounding landscape, especially inference
  - Teaching of lines should also relate to LOBFs
- Mathematics and Statistics
  - Awareness of disciplinary distinctions, particularly when more deterministic and more flexible approaches are appropriate in statistics

Thank you for coming!

Contact Stephanie Casey with comments/questions/good statistics jokes:

scasey1@emich.edu

Slides available at:

https://sites.google.com/site/stephaniecaseymath/