A Simplifying Framework for an Introductory Statistics Class

By
Dr. Mark Eakin
eakin@uta.edu
University of Texas at Arlington

Advance Organizer

- Instructional strategy to promote learning and retention or material used before instruction to help organize material that will be presented. Few if any technical terms used.
  http://advanceorganizers.wikispaces.com/All+About+Advance+Organizers
- Have been shown to work in many but not all studies (Meta-analysis article by C.L. Stone, 1960)
  http://www.jstor.org/stable/20151510

Random Rectangles

A set of 100 rectangles are displayed on one sheet of paper with the sizes of these rectangles being highly right-skewed (see handout). Students are asked to sample from these to illustrate sampling distributions

- Created by Dr. Richard Scheaffer and found in his book: *Activity Based Statistics*
- Numerous books now use versions of this
  http://www.gobookee.org/statistics-rectangle-activity/

Sampling Activity

Four Approaches to Estimate Average Size of All 100 Boxes

- A guess of the average size of the rectangles
- Students asked to randomly pick rectangles
- Students close their eyes and randomly point to an ID in a 10x10 table of rectangle ID numbers (Students did not know that I put all the large boxes IDs in the middle of the table.)
- Using their birth month and day students pick 10 (pseudo) random ID numbers from a table

Their four estimate are collected using Blackboard.

Results From One Semester*

- The results from 134 students are examined.
- The population is first described then the answers for each of the four estimation procedures
- First pass through the results focuses on the errors (sample mean – population mean) in each approach

	Guess	Student Random	Blind Point	Randon
Mean=	10.1	7.9359	8.969	7.034
St Dev=	4.25	3.6734	3.293	2.526
St. Err =	5.26	3.7932	3.841	2.526

- See Handout

First Building Block

After a discussion of the sizes of the errors and the biases, I give the first building block of the course:

"1. Random samples will be used because they tend to have smaller errors then other sampling approaches." (I do not talk about exceptions to this rule until later in the course.)

Second Building Block

After examining the errors, I ask if any of the samples means could be 6.99 which is impossible for the last three procedures since their divisor is 10. This leads to:

“2. Sample estimates tend not to equal the population value.”

Third Building Block

The errors are re-examined. We determine the percent of estimates that have an error of 5 or more in each case. The value of 5 is then compared to the standard error. From this discussion comes the third building block:

“3. To evaluate an error, compare it to the standard error.” (This is the foundation on which I build z and t tests)

Third Building Block Side-Note 1

To evaluate an error we examined the chance of that occurring by counting the number of times the error exceeded a value and forming a percent. This approach of creating multiple samples is time consuming and by examining the graphs of the distribution of sample means we see that it is approaching a bell curve. I state that this well known shape can be used later for this purpose.

Further in the semester, this sets up a discussion of the use of the z and t distributions to reduce the time it takes to calculating the chance that some value is unlikely.

Third Building Block Side-Note 2

Additionally, I mention that if the ratio in Building Block 3 is unlikely it could be the sample was unlikely or because the population mean has changed. I come back to this when starting hypothesis testing.

Third Building Block Side-Note 3

- A discussion is also started on the effect on the error of increasing the variation in the box sizes and the effect of sampling more than 10 boxes.
- From this discussion, I propose to the students that the standard error consists of two components: variability and knowledge; the foundation from which I later create the standard error formulas.

Fourth Building Block

The discussion is then brought back to errors but now focusing on the likely rather than the unlikely error values. We determine the percent of time an error will be in specified range (e.g., within ± 5). I then reword this into Building Block 4

“4. The margin of error is the largest error we expect with a specified probability.”

Fourth Building Block-Side Note 1

- Combining Building Block 3 Side-Note 1 with Building Block 4, I show how they are related.
- Building Block 3: $-\text{Value} \le \dfrac{\text{sample mean} - \text{population mean}}{\text{standard error}} \le \text{Value}$
- which leads to the margin of error
  $-\text{Value} * \text{S.E.} \le \text{sample mean} - \text{population mean} \le \text{Value} * \text{S.E.}$
- and by solving for the population mean is the foundation from which I build Confidence Intervals

Conclusion

- Using the Random Rectangles allows me to preview almost all topics in a first level statistics course with a single activity
- While I have not observed any noticeable improvements in grades, it has become much easier to relate one topic to another saving time for other active-learning activities.

A Simplifying Framework for an Introductory Statistics Class

By

Dr. Mark Eakin

eakin@uta.edu

University of Texas at Arlington

Advance Organizer

- Instructional strategy to promote learning and retention or material used before instruction to help organize material that will be presented. Few if any technical terms used.

  http://advanceorganizers.wikispaces.com/All+About+Advance+Organizers

- Have been shown to work in many but not all studies (Meta-analysis article by C.L. Stone, 1960)

  http://www.jstor.org/stable/20151510

Random Rectangles

A set of 100 rectangles are displayed on one sheet of paper with the sizes of these rectangles being highly right-skewed (see handout). Students are asked to sample from these to illustrate sampling distributions

- Created by Dr. Richard Scheaffer and found in his book: *Activity Based Statistics*
- Numerous books now use versions of this
  http://www.gobookee.org/statistics-rectangle-activity/

Sampling Activity

Four Approaches to Estimate Average Size of All 100 Boxes

- A guess of the average size of the rectangles
- Students asked to randomly pick rectangles
- Students close their eyes and randomly point to an ID in a 10x10 table of rectangle ID numbers (Students did not know that I put all the large boxes IDs in the middle of the table.)
- Using their birth month and day students pick 10 (pseudo) random ID numbers from a table

Their four estimate are collected using Blackboard.

Results From One Semester*

- The results from 134 students are examined.
- The population is first described then the answers for each of the four estimation procedures
- First pass through the results focuses on the errors (sample mean – population mean) in each approach

	Guess	Student Random	Blind Point	Randon
Mean=	10.1	7.9359	8.969	7.034
St Dev=	4.25	3.6734	3.293	2.526
St. Err =	5.26	3.7932	3.841	2.526

- See Handout

First Building Block

After a discussion of the sizes of the errors and the biases, I give the first building block of the course:

"1. Random samples will be used because they tend to have smaller errors then other sampling approaches." (I do not talk about exceptions to this rule until later in the course.)

Second Building Block

After examining the errors, I ask if any of the samples means could be 6.99 which is impossible for the last three procedures since their divisor is 10. This leads to:

“2. Sample estimates tend not to equal the population value.”

Third Building Block

The errors are re-examined. We determine the percent of estimates that have an error of 5 or more in each case. The value of 5 is then compared to the standard error. From this discussion comes the third building block:

“3. To evaluate an error, compare it to the standard error.” (This is the foundation on which I build z and t tests)

Third Building Block Side-Note 1

To evaluate an error we examined the chance of that occurring by counting the number of times the error exceeded a value and forming a percent. This approach of creating multiple samples is time consuming and by examining the graphs of the distribution of sample means we see that it is approaching a bell curve. I state that this well known shape can be used later for this purpose.

Further in the semester, this sets up a discussion of the use of the z and t distributions to reduce the time it takes to calculating the chance that some value is unlikely.

Third Building Block Side-Note 2

Additionally, I mention that if the ratio in Building Block 3 is unlikely it could be the sample was unlikely or because the population mean has changed. I come back to this when starting hypothesis testing.

Third Building Block Side-Note 3

- A discussion is also started on the effect on the error of increasing the variation in the box sizes and the effect of sampling more than 10 boxes.

- From this discussion, I propose to the students that the standard error consists of two components: variability and knowledge; the foundation from which I later create the standard error formulas.

Fourth Building Block

The discussion is then brought back to errors but now focusing on the likely rather than the unlikely error values. We determine the percent of time an error will be in specified range (e.g., within ± 5). I then reword this into Building Block 4

"4. The margin of error is the largest error we expect with a specified probability."

Fourth Building Block-Side Note 1

- Combining Building Block 3 Side-Note 1 with Building Block 4, I show how they are related.
- Building Block 3: $-\text{Value} \leq \dfrac{\text{sample mean} - \text{population mean}}{\text{standard error}} \leq \text{Value}$
- which leads to the margin of error

$$-\text{Value} * \text{S.E.} \leq \text{sample mean} - \text{population mean} \leq \text{Value} * \text{S.E.}$$

- and by solving for the population mean is the foundation from which I build Confidence Intervals

Conclusion

- Using the Random Rectangles allows me to preview almost all topics in a first level statistics course with a single activity

- While I have not observed any noticeable improvements in grades, it has become much easier to relate one topic to another saving time for other active-learning activities.