

A NEW LOOK AT SAMPLING DISTRIBUTION

ANDREI PERKHOUNKOV

ANDREI.PERKHOUNKOV@KIRKWOOD.EDU

 **Kirkwood**
COMMUNITY COLLEGE

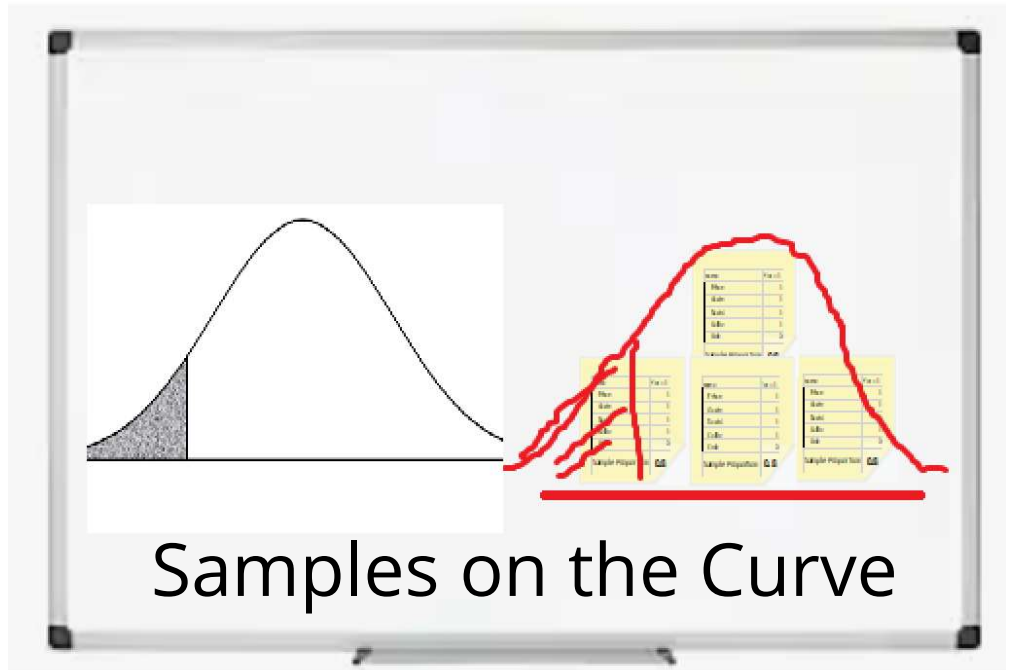
eCOTS 2024 - Beyond



	A	B
1	Student	Yes = 1
2	Rachael	1
3	Cooper	0
4	Caroline	1
5	Intisar	0
6	Makayla	1
7	Jayde	1
8	Christina	0
9	Dasol	1
10	Yasir	0
11	Jenna	1
12	Parker	0
13	Laura	0
14	Kyle	1
15	Grant	1
16	Khalid	0
17	Joanna	0
18	Noel	0
19	Nidhi	0
20	Van	1
21	Brisa	1
22	Fatima	0
23	Karly	0
24	Allison	0
25		
26	Proportion of "yes"	0.435
27		

	name	Yes = 1
4	Caroline	1
5	Intisar	0
8	Christina	0
15	Grant	1
16	Khalid	0
17	Joanna	0
19	Nidhi	0
22	Fatima	0
23	Karly	0
24	Allison	0
Sample Proportion		0.2

A sample may misrepresent the population.




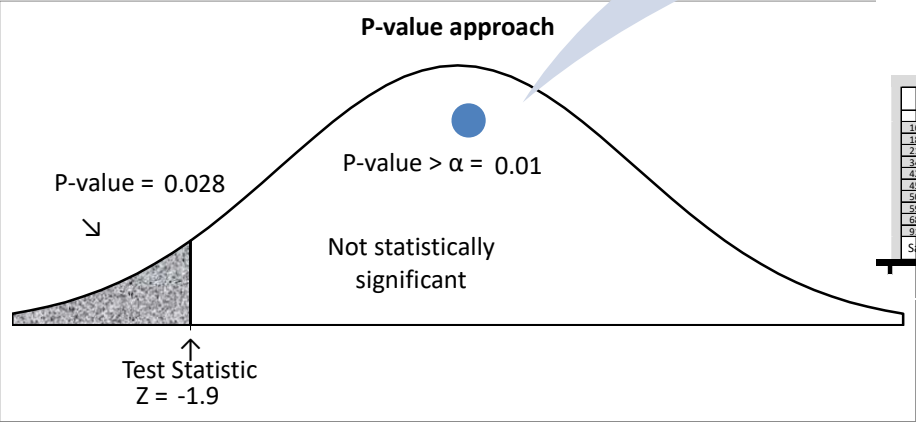
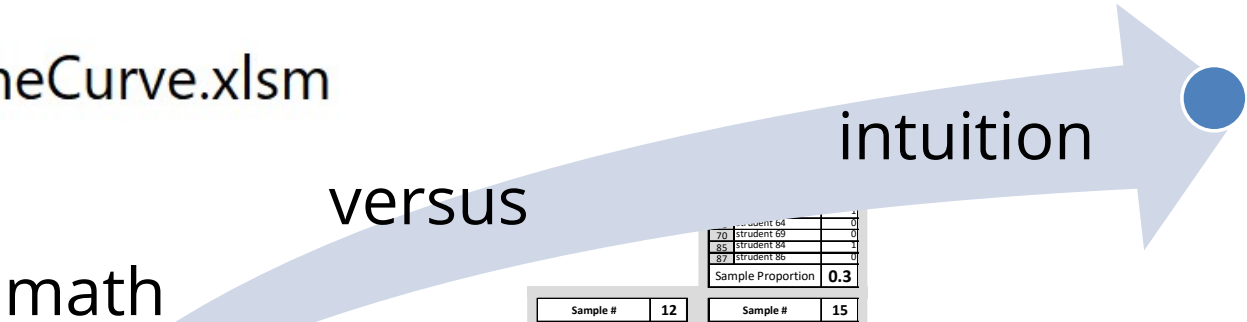
Samples on the Curve

P-value

Is the probability that a sample misrepresents the population.

Sampling Distribution as Samples on the Curve

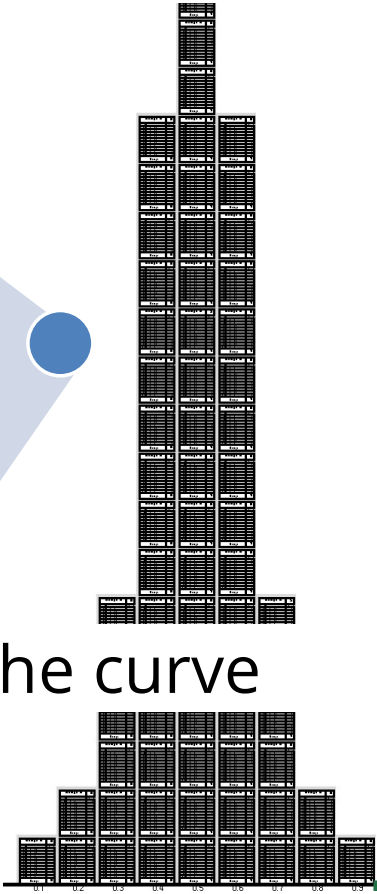
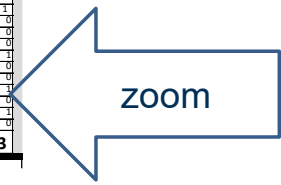
 SamplesOnTheCurve.xlsm



Sample #	12	15
name	Yes = 1	Yes = 1
8	strudent 7	10
17	strudent 16	17
20	strudent 19	26
24	strudent 22	28
35	strudent 34	35
37	strudent 36	61
57	strudent 56	85
67	strudent 66	90
94	strudent 93	97
97	strudent 96	94
Sample Proportion	0.2	Sample

Sample #	16	75	39
name	Yes = 1	Yes = 1	Yes = 1
16	strudent 15	8	5
18	strudent 17	11	17
21	strudent 20	24	20
34	strudent 33	35	28
42	strudent 41	44	46
45	strudent 44	67	53
50	strudent 49	61	61
59	strudent 58	75	76
68	strudent 67	90	85
91	strudent 90	99	87
Sample Proportion	0.1	Sample Proportion	0.3

Samples on the curve

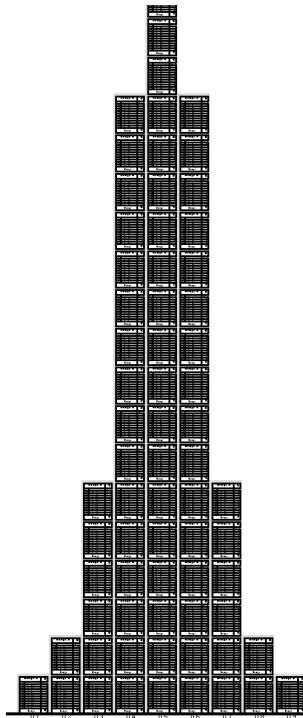


Samples on the Curve Tool Box?

https://drive.google.com/drive/folders/18Zn-m_QaLh11ATIVBmi-z5q16KoZH2e2?usp=sharing



SamplesOnTheCurve.xlsm



Excel macro-enabled workbook

- ✓ Point and click tool to get the Sampling Distribution of a sample proportion.
- ✓ Easy editing
- ✓ Artificial or real data

Example 1. Working with Classroom Data.

No prep time needed. It works great in a face-to-face or virtual classroom.

Ask students a question like:


- Are you driving or walking to class?
- Who watched the assigned video?
- What percentage of students finished module 1 homework?

Shoot for a question with close to 50% “yes” rate.

In my class, about 50% are driving, I will use this question:

Are you driving or walking to class?

Students may reply using google sheets, and I will copy/paste

to:  SamplesOnTheCurve.xlsx

Survey

Example 1. Working with Classroom Data.

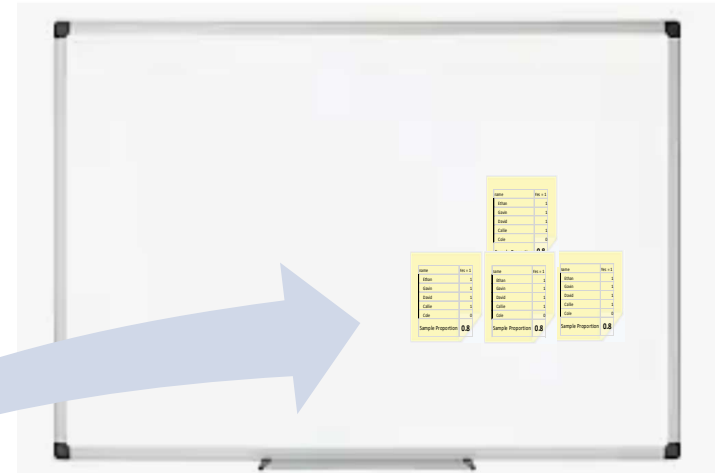
Survey: Are you driving to class?

Activity:

Students form teams of 5.
Write replies on a sticky note.

name	Yes = 1
Ethan	1
Gavin	1
David	1
Taylor	0
Callie	1

Sample Proportion	0.8
-------------------	------------

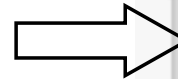


Teacher projects



SamplesOnTheCurve.xlsm

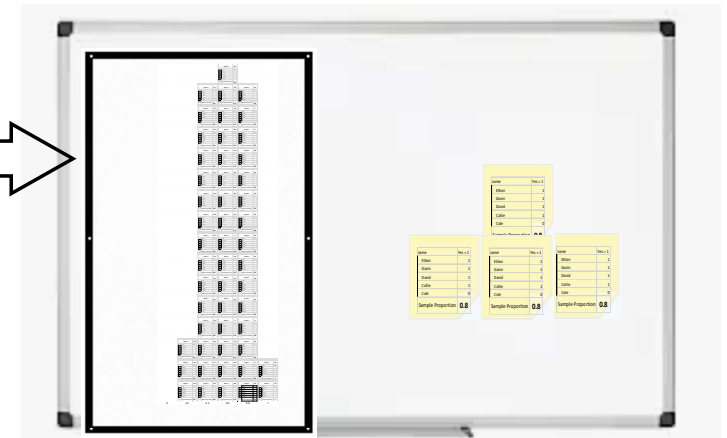
output



Discuss: 60% of students in the class are driving to class.

If you only asked 5 students, which is more likely?

16	Callie	1
17	Tristan	1
18	Alexis	1
19	Sydney	1
20	Bentley	1
21	Cole	0
22		
23	Proportion	0.6



- A. 2 out of 5 are driving
- B. 3 out of 5 are driving

Example 1. Working with Classroom Data.

Google Sheets is a good alternative to sticky notes, and works great in a web-live setting.

Hands-on or “mouse-on” activities help students internalize the notion of Standard Error.

“Group-to-group variability”

Vs.

“variability of sample statistic”

Mason	1
Ethan	1
Hannah	0
Daniel	0
Ellie	0
Eli	1
Gavin	1
Charlie	0
David	1
Isaiah	0
Chase	1
Jordan	1
Eliana	0
Taylor	0
Callie	1
Tristan	1
Alexis	1
Sydney	1
Bentley	1
Cole	0
Proportion	0.6

The image shows a grid of sample proportion tables. Each table represents a different sample size (Sample #) and a different sample proportion. The sample sizes range from 34 to 73. The sample proportions range from 0.1 to 0.5. A yellow sticky note is placed over one of the tables, showing a sample proportion of 0.8. The sticky note also lists the names of the students and their corresponding 'Yes' responses, which sum to 1.

Sample #	Sample Proportion
34	0.4
37	0.6
14	0.8
66	0.4
26	0.6
16	0.8
22	0.4
1	0.6
5	0.8
83	0.2
90	0.4
13	0.6
32	0.8
81	0.2
87	0.4
29	0.6
4	0.8
39	1
44	0.2
86	0.4
88	0.6
41	0.8
73	1

Example 1. Working with Classroom Data.

Are you driving to class?

Discuss:

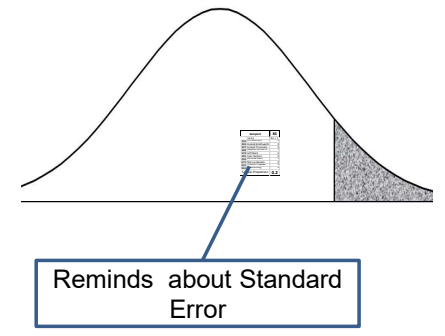
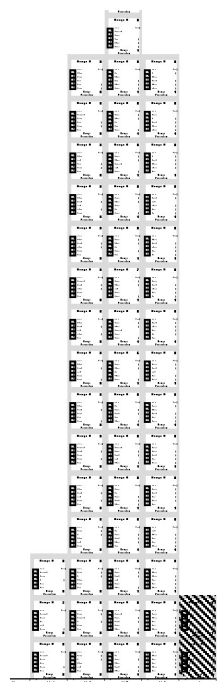
While 60% of students are driving, asking 5 students may misrepresent the overall proportion.

Counting samples, showing specific samples, shading samples on paper or Google Sheets, and forming groups based on specific samples depicted in the image—these and other activities proved to be efficient in explaining the area under the curve.

Mason	1
Ethan	1
Hannah	0
Daniel	0
Ellie	0
Eli	1
Gavin	1
Charlie	0
David	1
Isaiah	0
Chase	1
Jordan	1
Eliana	0
Taylor	0
Callie	1
Tristan	1
Alfred	1
Samuel	1
Matthew	1
Corey	0
Proportion	0.6

Sample #		79
name	Yes = 1	
2	Mason	1
8	Gavin	1
12	Chase	1
13	Jordan	1
16	Callie	1
Sample Proportion		1

Sample #		46
name	Yes = 1	
2	Mason	1
7	Eli	1
13	Jordan	1
16	Callie	1
17	Tristan	1
Sample Proportion		1

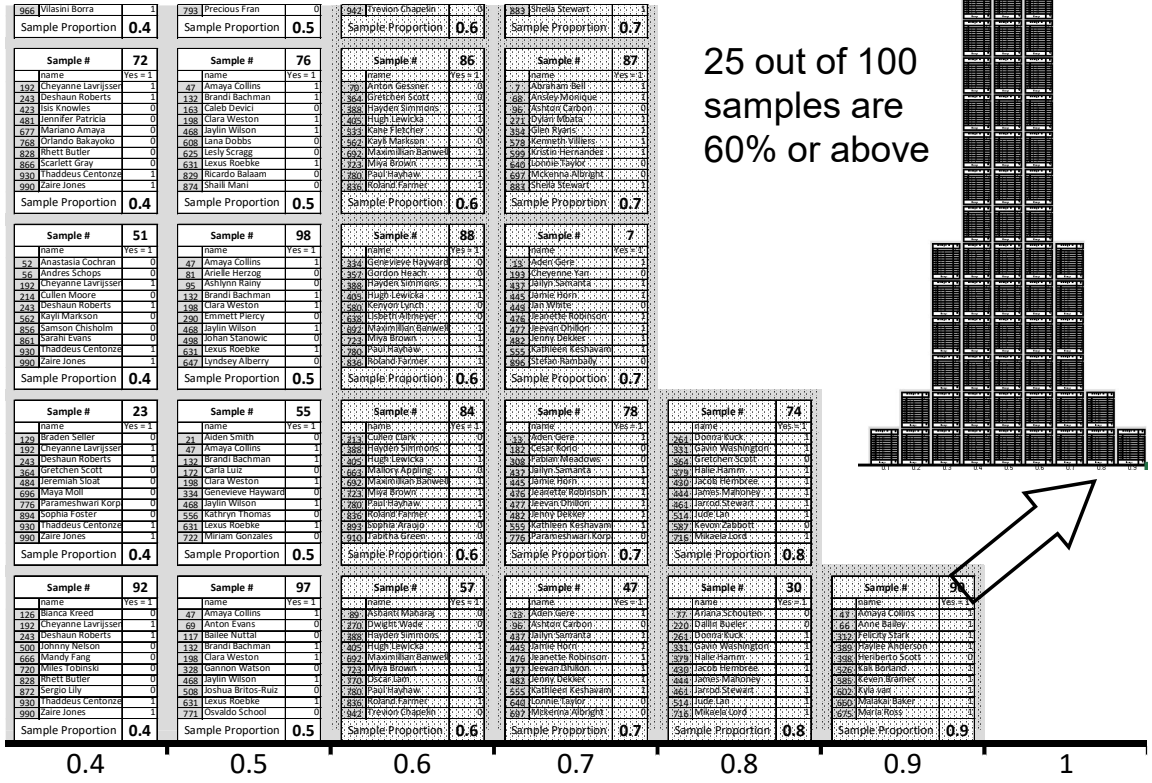


Example 2. Illustrating a word problem.

Discussion:

While overall 50% of respondents say 'yes,' a sample of size 10 may easily have a proportion of 'yes' of 60% or above.

	A	B
1	Respondent	Yes = 1
2	Aaliyah Fisher	1
3	Aaron Randall	0
4	Abby Cox	1
5	Abel Yeung	1
6	Abigale Diaz	1
7	Abraham Bell	1
8	Abraham Boxer	1
9	Abram Mccullough	1
10	Abril Daniels	1
11	Adeline Greene	1
12	Adeline Hendricks	0
13	Aden Gere	1
14	Adil Mutti	1
15	Aditya Wagenhaus	0
16	Adriel Lamain	1
17	Adrien Dozier	1
18	Adrianna Holland	1



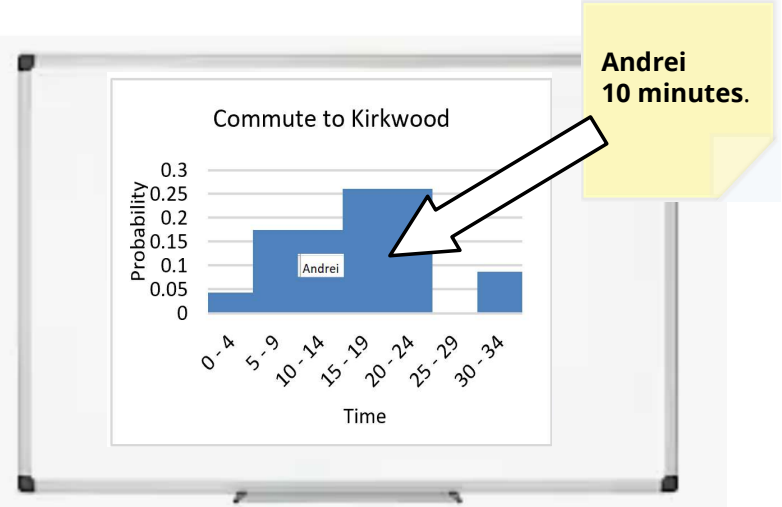
25 out of 100 samples are 60% or above

Pedagogy

If you would like to try 'Samples on the Curve,' start early in the course with activities that help students internalize what a histogram is. (I call it 'Individuals on the Curve,' and I also have a tool for that.)

I use classroom activities that involve creating histograms using

sticky notes for in-class sessions



or

Google Sheets for online classes.

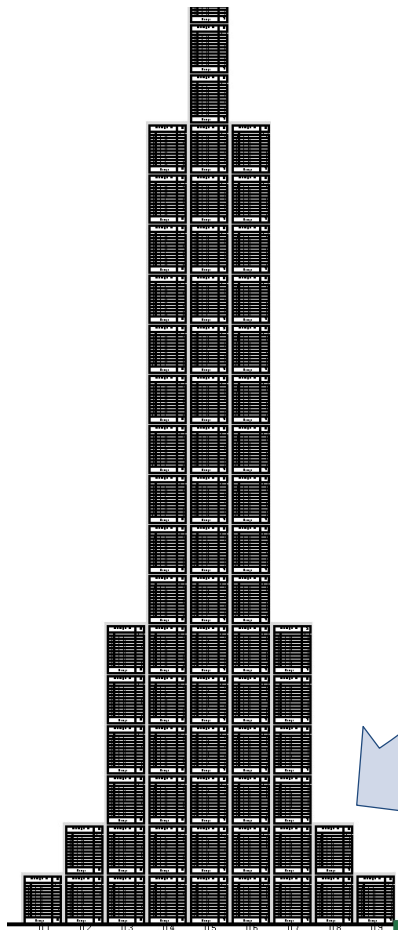
student first name	Time (minutes)					
Andrei	10					
Morgan	5					
Diana	17				Jaime	
Valeria	20				10	
Jaime	10				Katie	
Katie	10				10	
Gabby	15			Morgan	Andrei	Gabby
Jakub	10			5	10	15
Jacob	5					
Jordan	15					

Pedagogy

Samples on the Curve enable a student-friendly interpretation of the P-value.

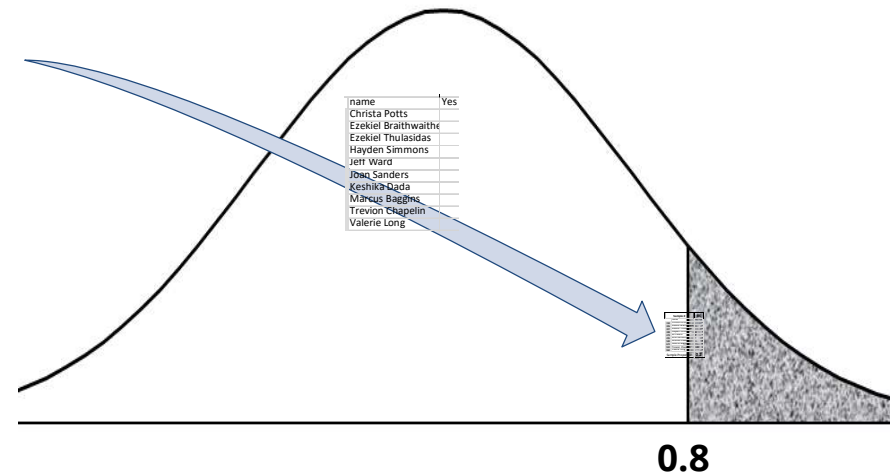
Interpretation of P-value:

The probability of getting a sample that misrepresents the population.



name	Yes = 1
Ethan	1
Gavin	1
David	1
Taylor	0
Allie	1

Sample Proportion **0.8**



Questions?