PSC 202
SYRACUSE UNIVERSITY

INTRODUCTION TO POLITICAL ANALYSIS

EXAM REVIEW
MORE SAMPLING AND SURVEYS

SURVEY

- Response rate: 87%
  - Extra credit for whole class

TODAY

- Exam review
- More on sampling and surveys

TODAY

- Exam review
- More on sampling and surveys

EXAM

- Wednesday: Exam #1
  - Bring a pen or pencil
  - Bring a calculator (no phone etc.)
  - Allowed to bring one single-page letter-size (8.5x11) sheet with you. Front page only. What you put on it is up to you, but it has to be your own sheet (we'll collect it)
- If you take exam at CDR, please sign up now!

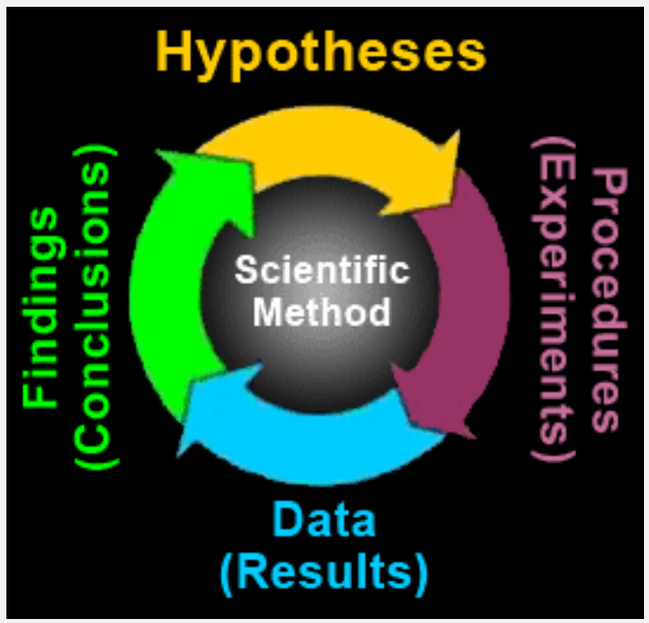
- Problem set 3 due on Friday
  - No quiz for Wednesday

EXAM

- Material covered
- Everything from August 28 (Course intro) to September 27 (Wednesday last week)

RESEARCH PROCESS

- Formulate research question
- Propose explanation/theory, hypotheses
- Data collection process
- **Use data to evaluate hypotheses**
- **Reassess explanation**



STUDY GUIDE

- Formulating a research question
  - What are characteristics of good/bad research questions?
- Be able to evaluate a research question
- Be able to come up with good research questions

STUDY GUIDE

- Developing a theory
  - What is the goal of a theory?
  - Dependent and independent variable
  - What are the characteristics of a good/bad theory?
  - Be able to evaluate a theory
  - Be able to come up with good theories

STUDY GUIDE

- Measurement process
  - What are the steps in the measurement process?
  - Conceptual definitions
    - concepts, unit of analysis, definition
    - be able to identify good/bad conceptual definitions, be able to come up with a good conceptual definition
- Operational definitions
  - be able to identify good/bad operational definitions, be able to come up with a good operational definition
- Measurement issues
  - reliability and validity

UNIT OF ANALYSIS

- **Major entity we want to analyze in a study: "unit of analysis"**
- **How do I identify what the unit of analysis is?**

UNIT OF ANALYSIS

- Look at dependent variable and independent variable

	Dependent Variable	Independent Variable
Observation 1		
Observation 2		
Observation 3		
Observation 4		
Observation 5		

UNIT OF ANALYSIS

- Look at dependent variable and independent variable

	Civil War	Economic Development: Gdp Per Capita
Observation 1		
Observation 2		
Observation 3		
Observation 4		
Observation 5		

UNIT OF ANALYSIS

- Look at dependent variable and independent variable

	Civil War	Economic Development: Gdp Per Capita
Observation 1	Yes	$3,000
Observation 2	No	$25,000
Observation 3	No	$4,000
Observation 4	Yes	$8,000
Observation 5	No	$45,000

UNIT OF ANALYSIS

- Look at dependent variable and independent variable

	Civil War	Economic Development: Gdp Per Capita
Observation 1	Yes	$3,000
Observation 2	No	$25,000
Observation 3	No	$4,000
Observation 4	Yes	$8,000
Observation 5	No	$45,000

- Ask: Who or what has civil wars (or not), and who or what has different levels of GDP per capita?

UNIT OF ANALYSIS

- Look at dependent variable and independent variable

	Civil War	Economic Development: Gdp Per Capita
Bob	Yes	$3,000
Mary	No	$25,000
Gene	No	$4,000
Anna	Yes	$8,000
Emma	No	$45,000

- People? Nope…

UNIT OF ANALYSIS

- Look at dependent variable and independent variable

	Civil War	Economic Development: Gdp Per Capita
Seattle	Yes	$3,000
NYC	No	$25,000
LA	No	$4,000
Chicago	Yes	$8,000
Miami	No	$45,000

- Cities? Nope…

UNIT OF ANALYSIS

- Look at dependent variable and independent variable

	Civil War	Economic Development: Gdp Per Capita
Venezuela	Yes	$3,000
Canada	No	$25,000
Botswana	No	$4,000
Greenland	Yes	$8,000
USA	No	$45,000

- Countries? Much better

Values for variables are made up!

UNIT OF ANALYSIS

- Look at dependent variable and independent variable

	Civil War	Economic Development: Gdp Per Capita
Venezuela	Yes	$3,000
Canada	No	$25,000
Botswana	No	$4,000
Greenland	Yes	$8,000
USA	No	$45,000

- Unit of Analysis: Countries

UNIT OF ANALYSIS

- Another example

	Rating Of D. Trump	Gender
Observation 1	56	Male
Observation 2	34	Male
Observation 3	67	Female
Observation 4	10	Female
Observation 5	97	Male

- Unit of Analysis: ?

UNIT OF ANALYSIS

- Another example

	Rating Of D. Trump	Gender
Bob	56	Male
Gene	34	Male
Anna	67	Female
Mary	10	Female
Eric	97	Male

- Unit of Analysis: Individuals

UNIT OF ANALYSIS

- **Major entity we want to analyze in a study: "unit of analysis"**
  - **The dependent and independent variable have the same unit of analysis!**

STUDY GUIDE

- Variables
  - Variable labels and values
  - Measurement levels
    - nominal, ordinal, interval
- Central tendency
  - mode, median, mean
- Dispersion
  - interquartile range, standard deviation

CENTRAL TENDENCY

- Student grades during freshman year (ordered from best to worst)

	Grade
Class 1	A
Class 2	A
Class 3	A-
Class 4	A-
Class 5	A-
Class 6	B+
Class 7	B
Class 8	B
Class 9	B-
Class 10	C
Class 11	C

- What central tendency can we compute?

CENTRAL TENDENCY

- **Median?**

	Grade
Class 1	A
Class 2	A
Class 3	A-
Class 4	A-
Class 5	A-
Class 6	B+
Class 7	B
Class 8	B
Class 9	B-
Class 10	C
Class 11	C

CENTRAL TENDENCY

- **Median?**

	Grade
Class 1	A
Class 2	A
Class 3	A-
Class 4	A-
Class 5	A-
Class 6	B+
Class 7	B
Class 8	B
Class 9	B-
Class 10	C
Class 11	C

Student did better in 5 classes

Right in middle

Student did worse in 5 classes

FREQUENCY TABLE

Grade	Number	Percentage	Cumulative Percentage
A	2		
A-	3		
B+	1		
B	2		
B-	1		
C	2		

- Same information, presented differently

FREQUENCY TABLE

Grade	Number	Percentage	Cumulative Percentage
A	2	(2/11)*100=18.2%	
A-	3		
B+	1		
B	2		
B-	1		
C	2		

- Same information, presented differently

FREQUENCY TABLE

Grade	Number	Percentage	Cumulative Percentage
A	2	(2/11)*100=18.2%	
A-	3	(3/11)*100=27.3%	
B+	1	(1/11)*100=9.1%	
B	2	(2/11)*100=18.2%	
B-	1	(1/11)*100=9.1%	
C	2	(2/11)*100=18.2%	

- Same information, presented differently

FREQUENCY TABLE

Grade	Number	Percentage	Cumulative Percentage
A	2	18.2%	
A-	3	27.3%	
B+	1	9.1%	
B	2	18.2%	
B-	1	9.1%	
C	2	18.2%	

- Same information, presented differently

FREQUENCY TABLE

Grade	Number	Percentage	Cumulative Percentage
A	2	18.2%	18.2%
A-	3	27.3%	
B+	1	9.1%	
B	2	18.2%	
B-	1	9.1%	
C	2	18.2%	

- Same information, presented differently

FREQUENCY TABLE

Grade	Number	Percentage	Cumulative Percentage
A	2	18.2%	18.2%
A-	3	27.3%	18.2% + 27.3% = 45.5%
B+	1	9.1%	
B	2	18.2%	
B-	1	9.1%	
C	2	18.2%	

- Same information, presented differently

FREQUENCY TABLE

Grade	Number	Percentage	Cumulative Percentage
A	2	18.2%	18.2%
A-	3	27.3%	18.2% + 27.3% = 45.5%
B+	1	9.1%	45.5% + 9.1% = 54.6%
B	2	18.2%	
B-	1	9.1%	
C	2	18.2%	

- Same information, presented differently

FREQUENCY TABLE

Grade	Number	Percentage	Cumulative Percentage
A	2	18.2%	18.2%
A-	3	27.3%	18.2% + 27.3% = 45.5%
B+	1	9.1%	45.5% + 9.1% = 54.6%
B	2	18.2%	54.6% + 18.2% = 72.8%
B-	1	9.1%	72.8% + 9.1% = 81.9%
C	2	18.2%	81.9% + 18.2.1% = 100.1%

- Same information, presented differently

FREQUENCY TABLE

Grade	Number	Percentage	Cumulative Percentage
A	2	18.2%	18.2%
A-	3	27.3%	45.5%
B+	1	9.1%	54.6%
B	2	18.2%	72.8%
B-	1	9.1%	81.9%
C	2	18.2%	100.1%

- Median?

FREQUENCY TABLE

18.2% are an A—does not include 50%

Grade	Number	Percentage	Cumulative Percentage
A	2	18.2%	18.2%
A-	3	27.3%	45.5%
B+	1	9.1%	54.6%
B	2	18.2%	72.8%
B-	1	9.1%	81.9%
C	2	18.2%	100.1%

?

- We’re looking for the category whose cumulative percentage includes 50%

FREQUENCY TABLE

45.5% are an A- or better—does not include 50%

Grade	Number	Percentage	Cumulative Percentage
A	2	18.2%	18.2%
A-	3	27.3%	45.5%
B+	1	9.1%	54.6%
B	2	18.2%	72.8%
B-	1	9.1%	81.9%
C	2	18.2%	100.1%

?

- We’re looking for the category whose cumulative percentage includes 50%

FREQUENCY TABLE

54.6% are an B+ or better

Grade	Number	Percentage	Cumulative Percentage
A	2	18.2%	18.2%
A-	3	27.3%	45.5%
B+	1	9.1%	54.6% !
B	2	18.2%	72.8%
B-	1	9.1%	81.9%
C	2	18.2%	100.1%

- Cumulative percentage jumps from under 50% to over 50%

FREQUENCY TABLE

Grade	Number	Percentage	Cumulative Percentage
A	2	18.2%	18.2%
A-	3	27.3%	45.5%
B+	1	9.1%	54.6%
B	2	18.2%	72.8%
B-	1	9.1%	81.9%
C	2	18.2%	100.1%

- Median grade is B+ [Median grade is NOT 1 or 9.1%]

MEDIAN

- Variables
  - Variable labels and values
  - Measurement levels
    - nominal, ordinal, interval
- Central tendency
  - mode, median, mean
- Dispersion
  - interquartile range, standard deviation

CENTRAL TENDENCY

- **Mean?**

	Grade
Class 1	A
Class 2	A
Class 3	A-
Class 4	A-
Class 5	A-
Class 6	B+
Class 7	B
Class 8	B
Class 9	B-
Class 10	C
Class 11	C

CENTRAL TENDENCY

- **Mean?**

	Grade	Grade Point Equivalent
Class 1	A	4.0
Class 2	A	4.0
Class 3	A-	3.7
Class 4	A-	3.7
Class 5	A-	3.7
Class 6	B+	3.3
Class 7	B	3.0
Class 8	B	3.0
Class 9	B-	2.7
Class 10	C	2.0
Class 11	C	2.0

CENTRAL TENDENCY

- **Mean?**

	Grade	Grade Point Equivalent
Class 1	A	4.0
Class 2	A	4.0
Class 3	A-	3.7
Class 4	A-	3.7
Class 5	A-	3.7
Class 6	B+	3.3
Class 7	B	3.0
Class 8	B	3.0
Class 9	B-	2.7
Class 10	C	2.0
Class 11	C	2.0

- 4.0 + 4.0 + 3.7 + 3.7 + 3.7 + 3.3 + 3.0 + 3.0 + 2.7 + 2.0 + 2.0 = 35.1
- 35.1 / 11 = 3.19

FREQUENCY TABLE

Grade	Grade Point Equivalent	Number	Percentage
A	4.0	2	18.2%
A-	3.7	3	27.3%
B+	3.3	1	9.1%
B	3.0	2	18.2%
B-	2.7	1	9.1%
C	2.0	2	18.2%

- Same information, presented differently

FREQUENCY TABLE

Grade	Grade Point Equivalent	Number	Percentage
A	4.0	2	18.2%
A-	3.7	3	27.3%
B+	3.3	1	9.1%
B	3.0	2	18.2%
B-	2.7	1	9.1%
C	2.0	2	18.2%

- (2 x 4.0)

FREQUENCY TABLE

Grade	Grade Point Equivalent	Number	Percentage
A	4.0	2	18.2%
A-	3.7	3	27.3%
B+	3.3	1	9.1%
B	3.0	2	18.2%
B-	2.7	1	9.1%
C	2.0	2	18.2%

- (2 x 4.0) + (3 x 3.7)

FREQUENCY TABLE

Grade	Grade Point Equivalent	Number	Percentage
A	4.0	2	18.2%
A-	3.7	3	27.3%
B+	3.3	1	9.1%
B	3.0	2	18.2%
B-	2.7	1	9.1%
C	2.0	2	18.2%

- (2 x 4.0) + (3 x 3.7) + (1 x 3.3)

FREQUENCY TABLE

Grade	Grade Point Equivalent	Number	Percentage
A	4.0	2	18.2%
A-	3.7	3	27.3%
B+	3.3	1	9.1%
B	3.0	2	18.2%
B-	2.7	1	9.1%
C	2.0	2	18.2%

- (2 x 4.0) + (3 x 3.7) + (1 x 3.3) + (2 x 3.0)

FREQUENCY TABLE

Grade	Grade Point Equivalent	Number	Percentage
A	4.0	2	18.2%
A-	3.7	3	27.3%
B+	3.3	1	9.1%
B	3.0	2	18.2%
B-	2.7	1	9.1%
C	2.0	2	18.2%

- (2 x 4.0) + (3 x 3.7) + (1 x 3.3) + (2 x 3.0) + (1 x 2.7)

FREQUENCY TABLE

Grade	Grade Point Equivalent	Number	Percentage
A	4.0	2	18.2%
A-	3.7	3	27.3%
B+	3.3	1	9.1%
B	3.0	2	18.2%
B-	2.7	1	9.1%
C	2.0	2	18.2%

- (2 x 4.0) + (3 x 3.7) + (1 x 3.3) + (2 x 3.0) + (1 x 2.7) + (2 x 2.0) = 35.1

FREQUENCY TABLE

Grade	Grade Point Equivalent	Number	Percentage
A	4.0	2	18.2%
A-	3.7	3	27.3%
B+	3.3	1	9.1%
B	3.0	2	18.2%
B-	2.7	1	9.1%
C	2.0	2	18.2%

- (2 x 4.0) + (3 x 3.7) + (1 x 3.3) + (2 x 3.0) + (1 x 2.7) + (2 x 2.0) = 35.1
- 35.1 / 11 = 3.19

STUDY GUIDE

- Sampling
  - Will not be on this exam

EXAM

- Important in exam: Show your steps!

EXAM

- Important in exam: Time management!
- Time spent on each question should be proportional to how many points they are

EXAM

- What to do and bring
  - Be a few minutes early
  - Bring calculator and pen/pencil
  - Bring your cheat sheet
  - Show your steps
  - Remember time management
  - Write legibly
  - If something is unclear, ask!

EXAM

- Questions?

TODAY

- Exam review
- More on sampling and surveys

RECAP

POLITICS JANUARY 25, 2023

Biden Averaged 41% Job Approval in His Second Year

Results for this Gallup poll are based on telephone interviews conducted Jan. 2-22, 2023, with a random sample of 1,011 adults, aged 18 and older, living in all 50 U.S. states and the District of Columbia. For results based on the total sample of national adults, the margin of sampling error is ±4 percentage points at the 95% confidence level. All reported margins of sampling error include computed design effects for weighting.

RECAP

- Population: the entire universe of objects to which our hypothesis applies
- Sample: the subset of the population that we study in order to make inferences about the full population

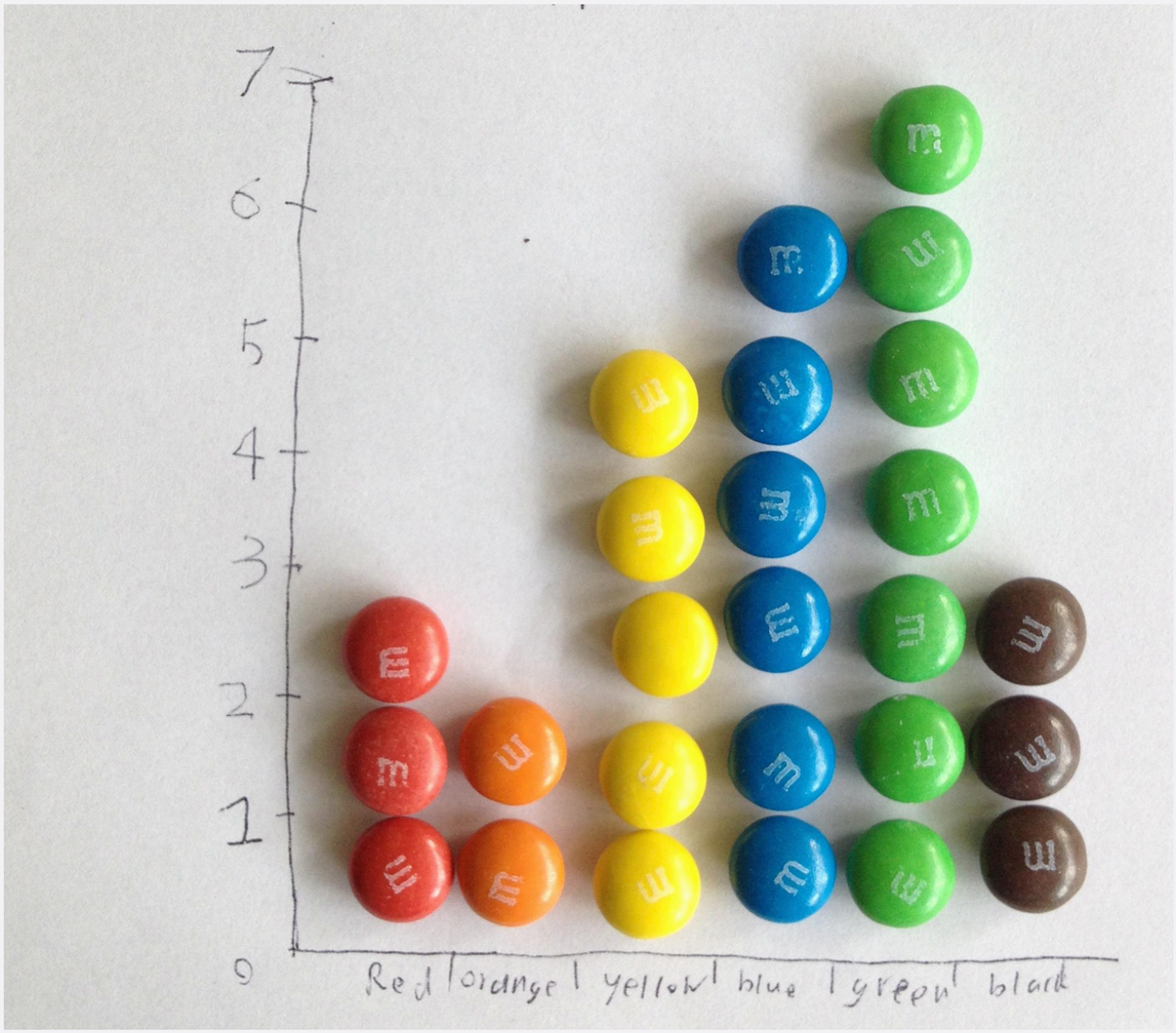
- We often use random sample of the population to learn about the population

RANDOM SAMPLING

- A random sample of the population avoids *systematic* sampling error
  - We will not get a sample that systematically over-represents certain characteristics of the population
    - e.g. random sample will not be 80% men and 20% women
    - Or mostly rich people
- If we use random sampling, we can use our sample's characteristics to estimate the population's characteristics
  - e.g. can use 1,011 randomly selected survey respondents to infer approval rating of J. Biden in American population

RANDOM SAMPLING ERROR

- But: random sampling introduces *random* sampling error



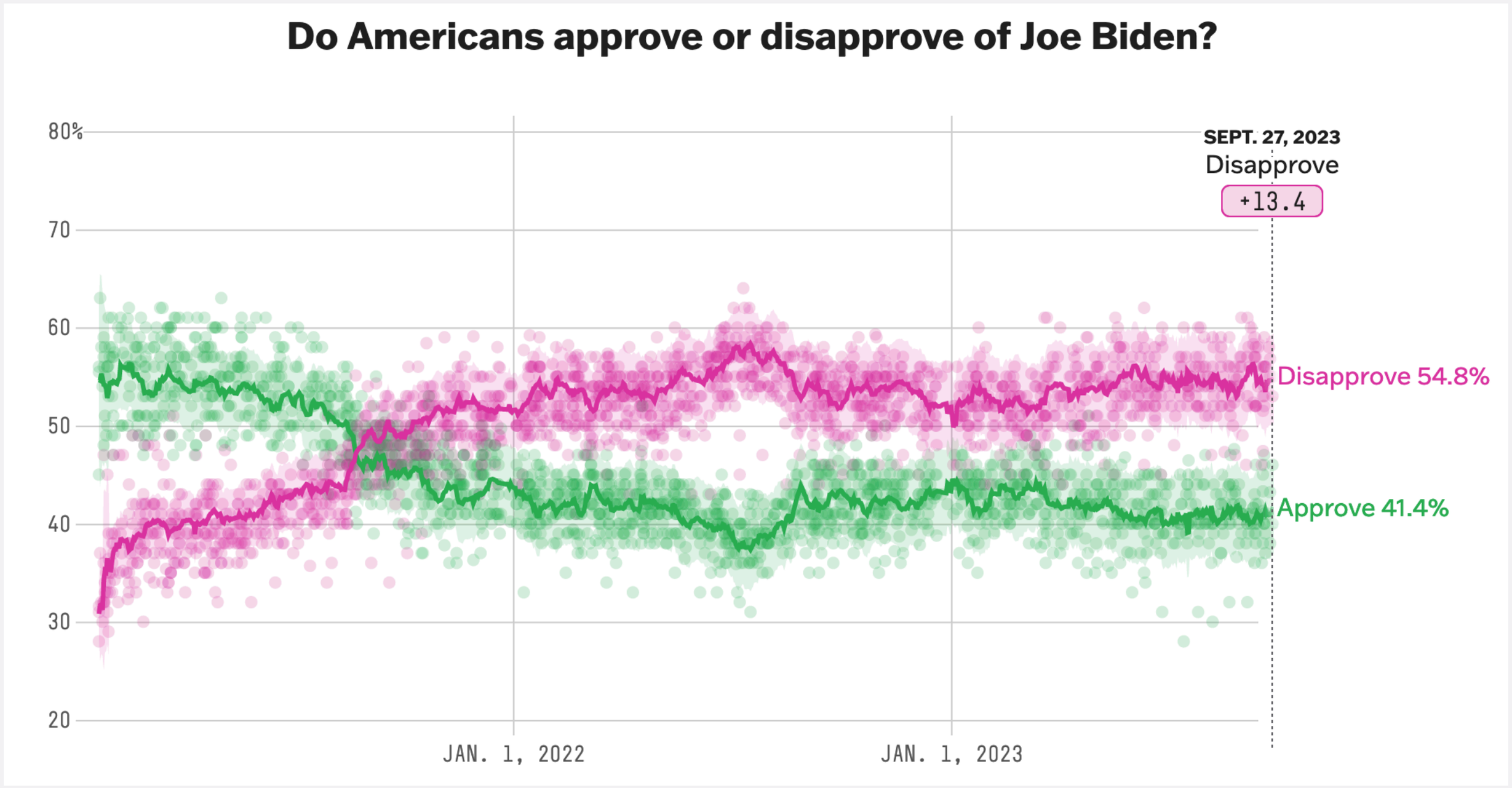
RANDOM SAMPLING ERROR

- Random sampling introduces *random* sampling error
  - Rolling a die
  - For a fair dice, we know that 1 to 6 occur with equal probability
- We roll a dice 6 times:
  - We may get each number exactly once
  - May also get 6 five times and 5 once
  - Or just 1s and 2s.
  - Without this random sampling error, many board games would not be a lot of fun

RANDOM SAMPLING ERROR

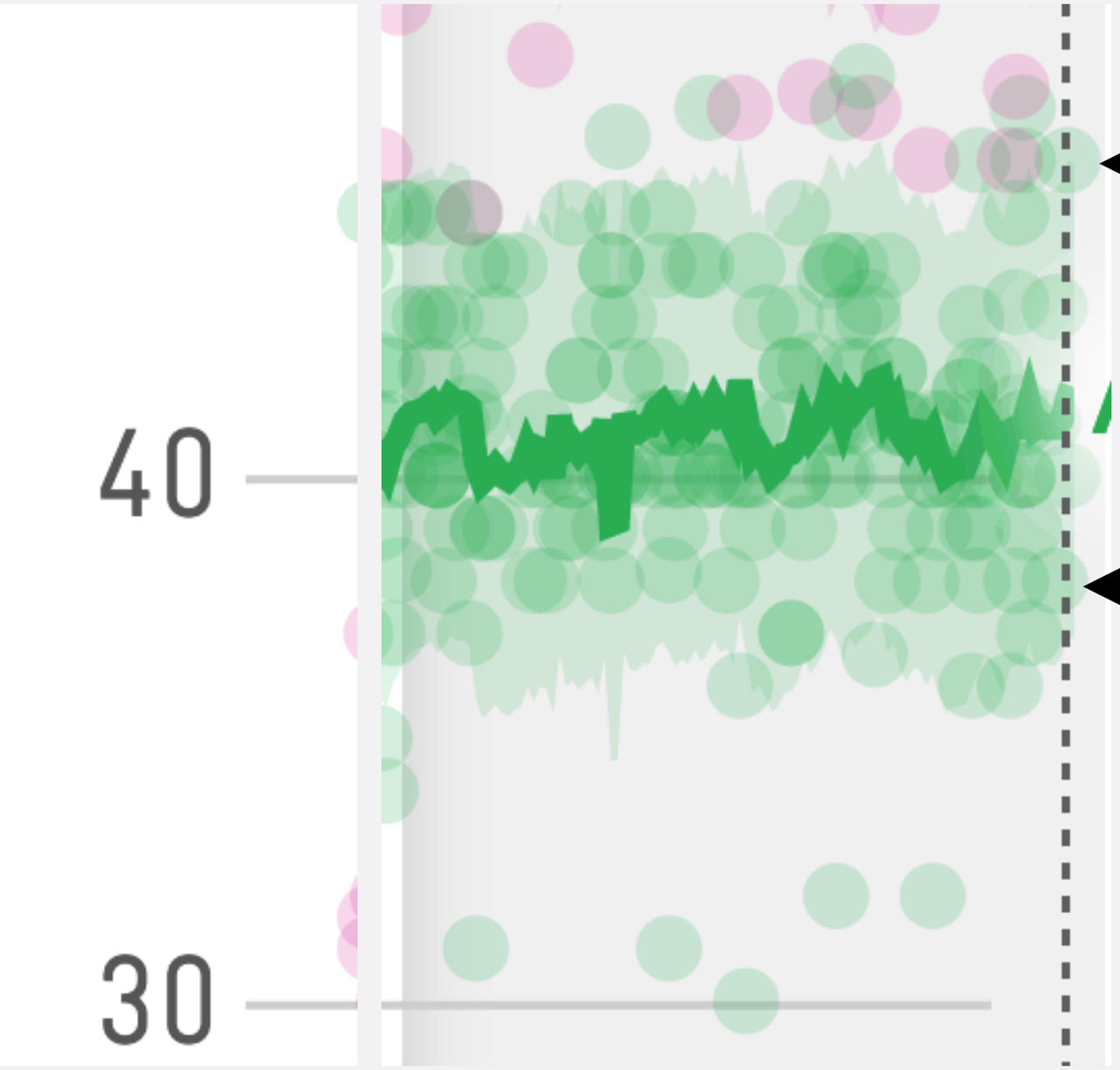
- Same logic when drawing sample from population
- It is unlikely that our random sample looks *exactly* like the population
  - e.g. by chance, we might draw more people that approve of Biden than is the case in the population
  - Or we might draw more people that disapprove of his performance than in the population
  - Problem: We don't know whether we've drawn a sample with fewer or more Biden supporters than in the population

RANDOM SAMPLING ERROR



- https://projects.fivethirtyeight.com/polls/approval/joe-biden/

RANDOM SAMPLING ERROR



- Which sample is closer to the population? We don’t know!

THE PROBLEM

Unknown: Approval rating in population

Known: Approval rating in survey

- Population parameter = Sample statistic + random sampling error

Also unknown

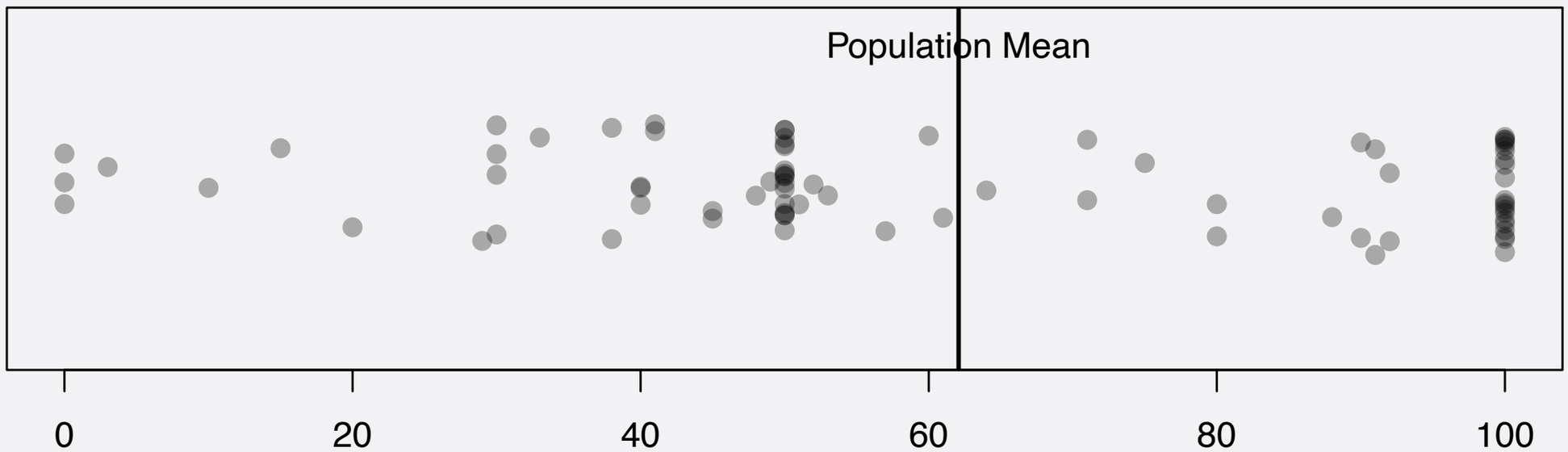
GOOD NEWS

- We can figure out how large the random sampling error is

OUR SURVEY

- **Feelings towards Taylor Swift among population of PSC 202 students**
- **On a scale from 0 to 100**

OUR SURVEY



- **Mean: 62.1**
- **Standard deviation: 30.0**

POPULATION

- This is the view of among the *population* of PSC 202 students
  - Usually: survey of sample, not of population
  - Here: we *do* have the population
  - What we can do: See what would happen if we only had a random sample of PSC 202 students, and compare that to the population
  - This will give us an idea of how large the sampling error is

THE PROBLEM

Here we know approval rating in population

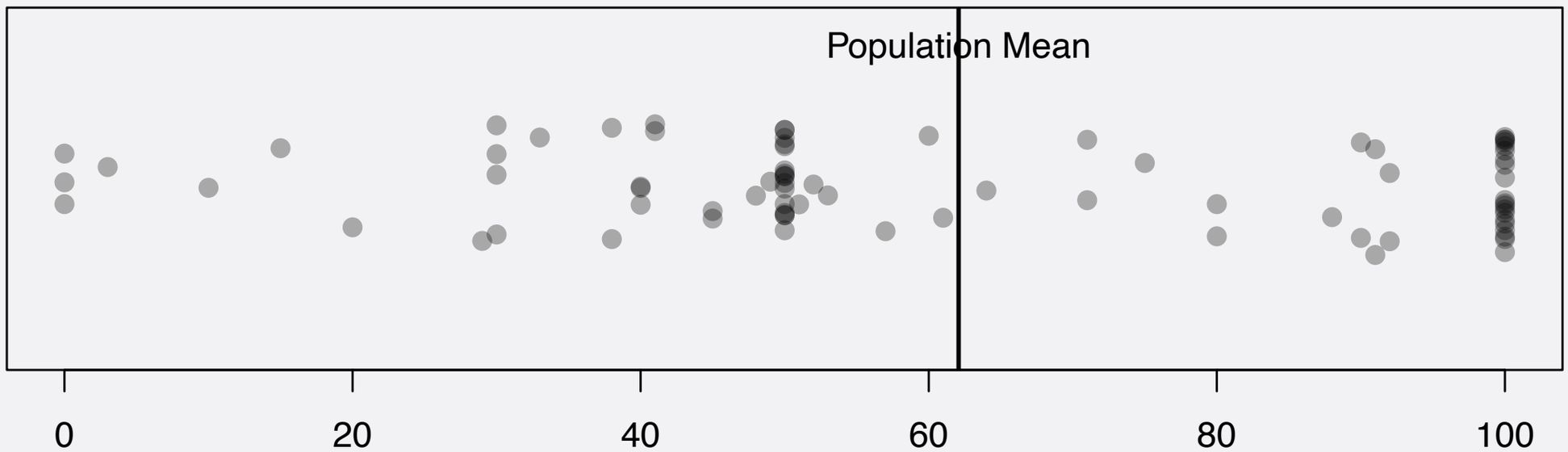
Known: Approval rating in survey

- Population parameter = Sample statistic + random sampling error

Unknown

But we can figure this out now

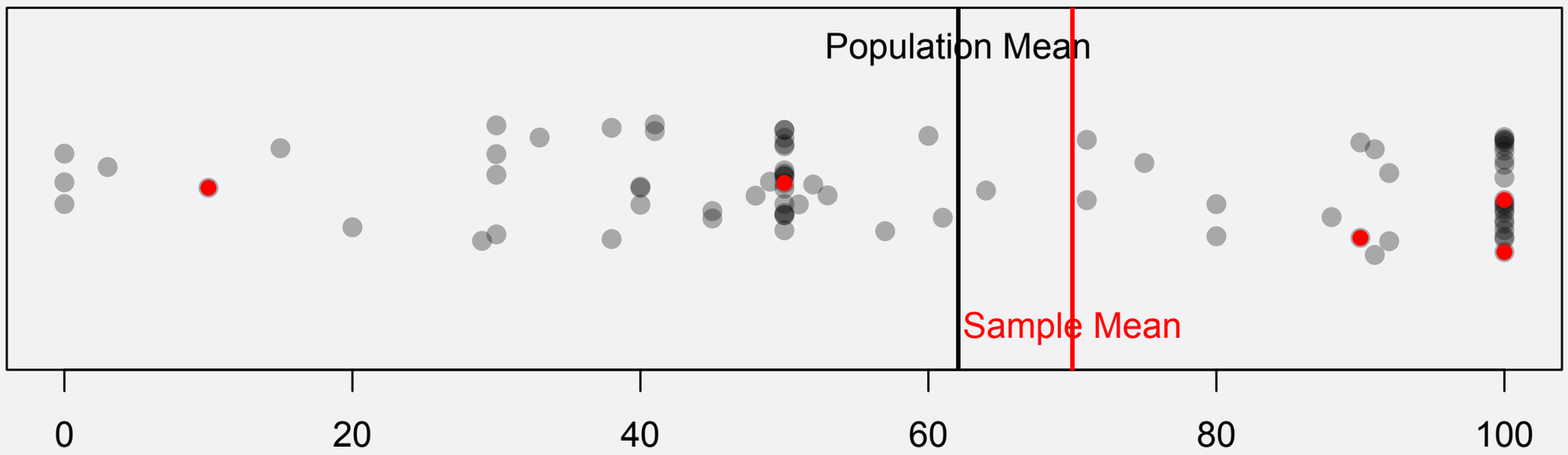
OUR SURVEY



- Mean: 62.1
- Standard deviation: 30.0

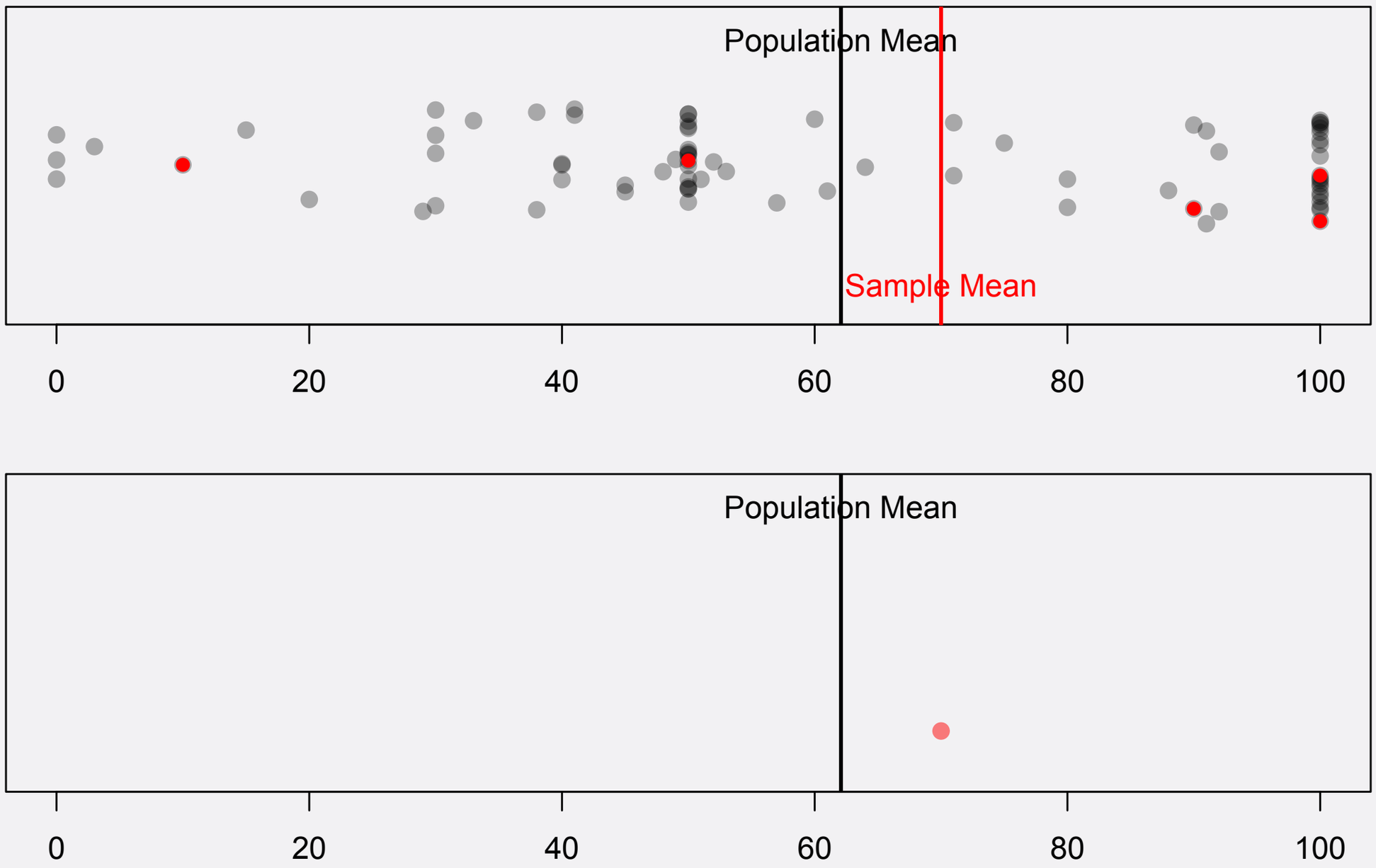
- Let’s say we could only survey a random sample of 5 students of PSC 202

OUR SURVEY

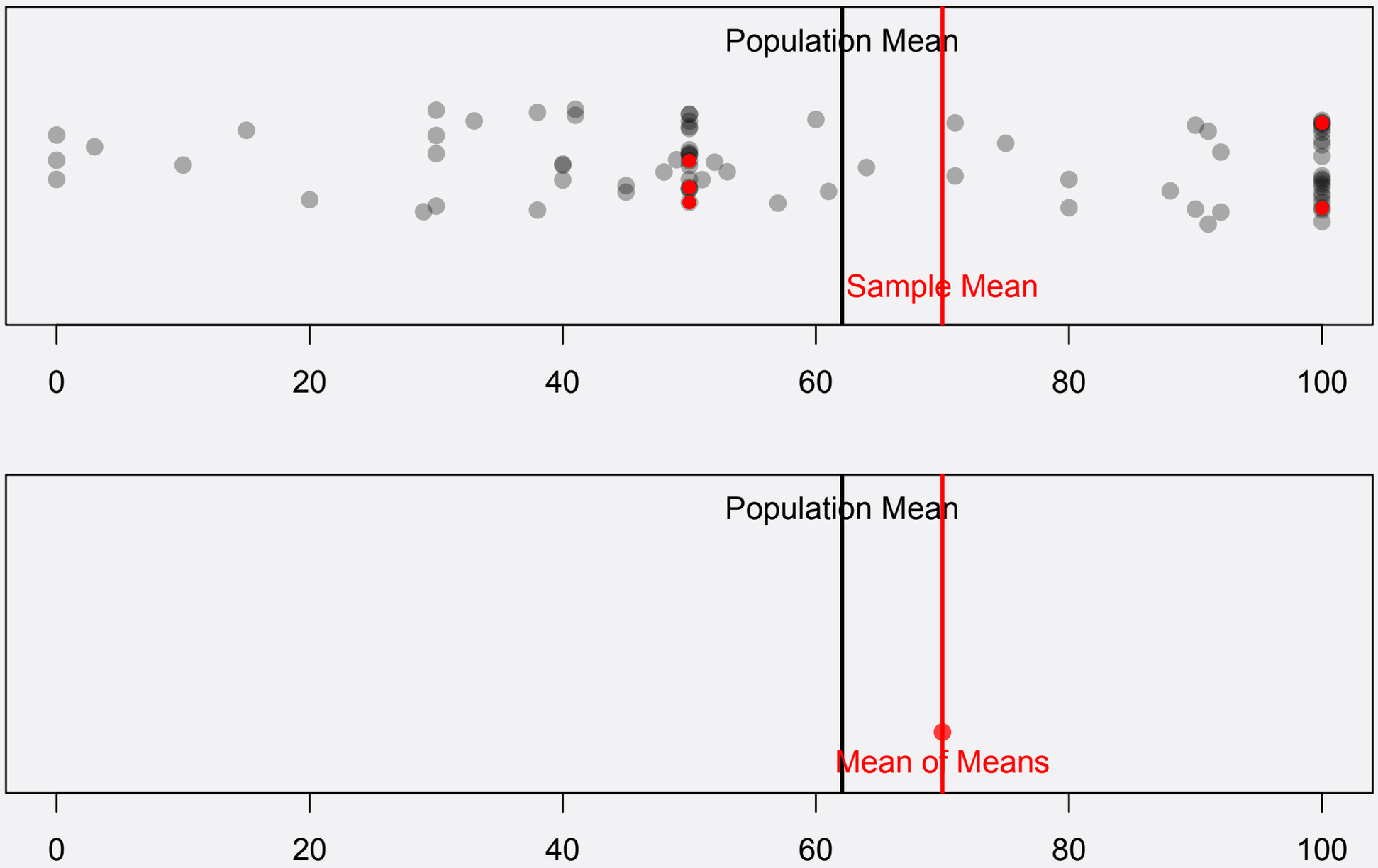


- **Mean of this sample: higher than the population mean**

OUR SURVEY

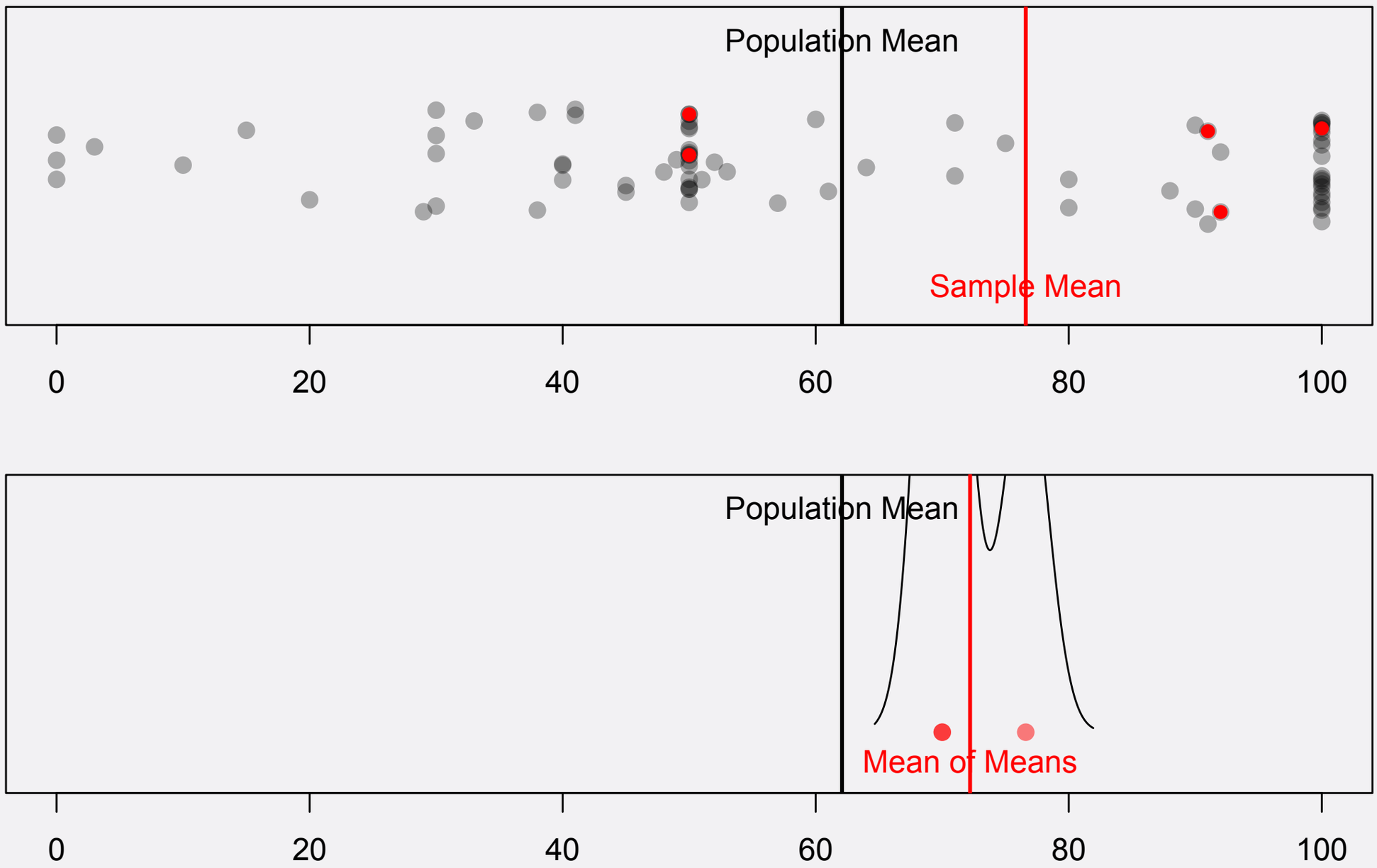

Population Mean
Sample Mean
0
20
40
60
80
100
Population Mean
0
20
40
60
80
100

OUR SURVEY



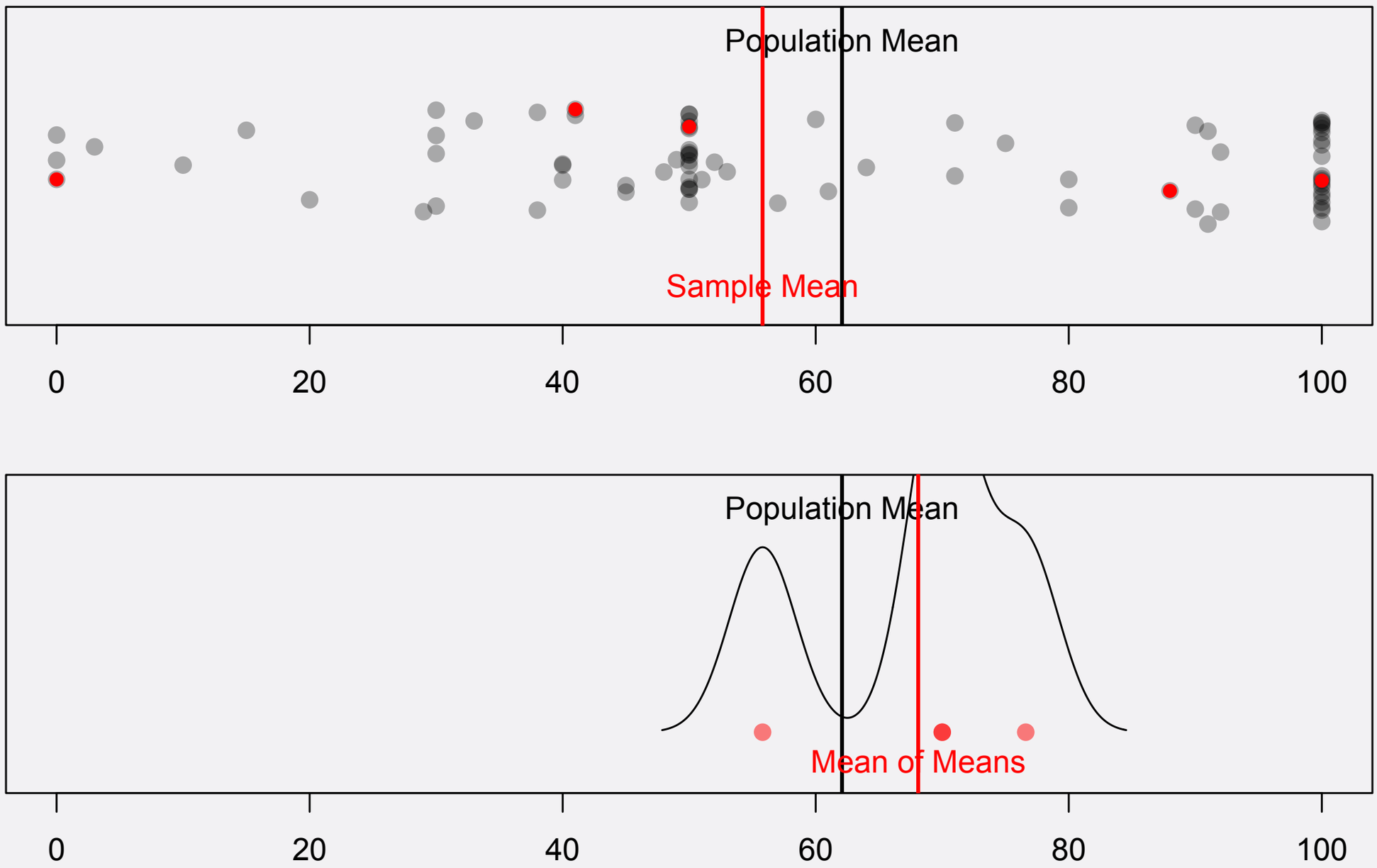
- **Let's do this again…another random sample of 5**
- **The mean here is again higher than in the population!**

OUR SURVEY



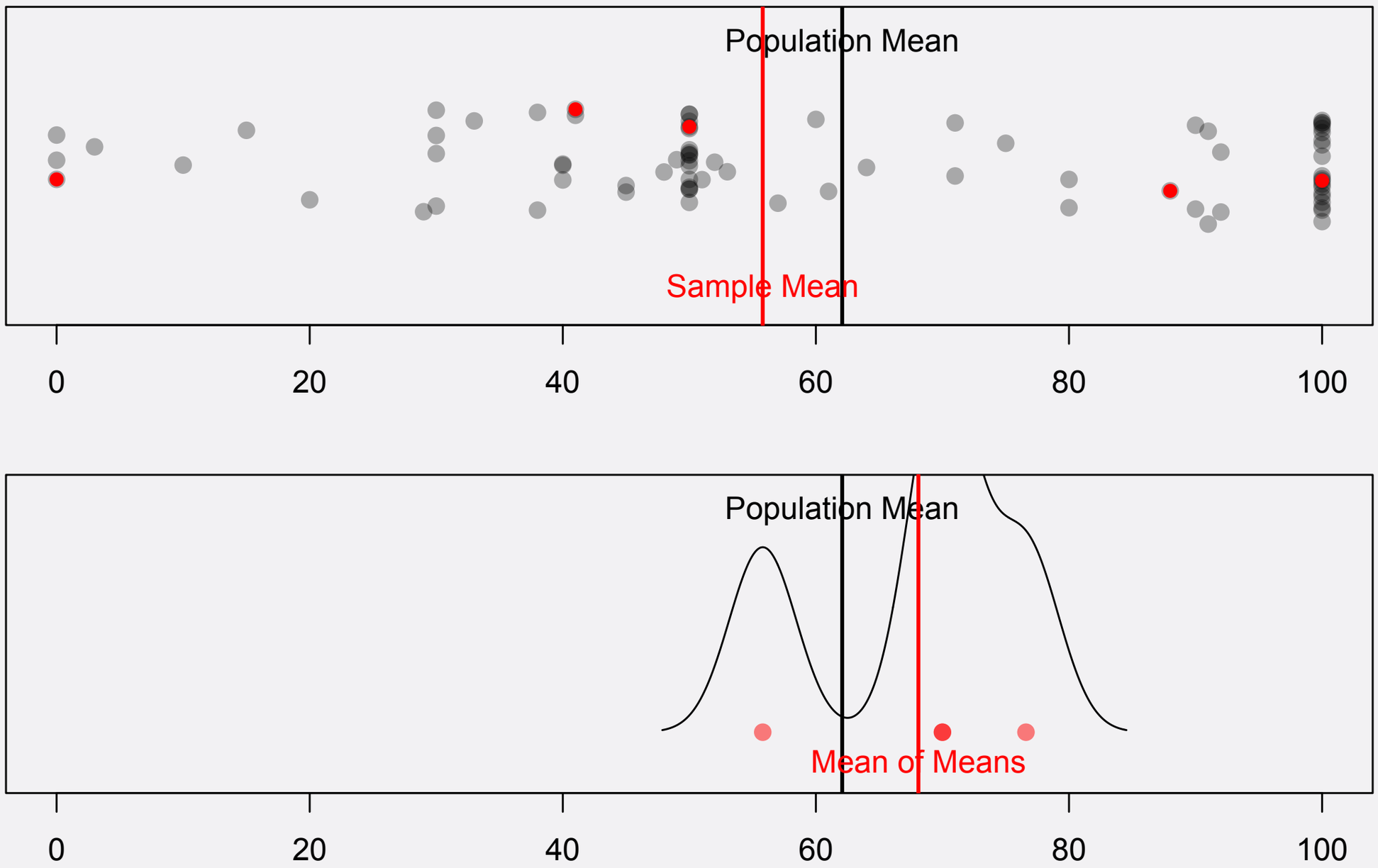
- A third random sample of 5: again higher!
- So we have three polls that estimate higher approval than is the case in the population

OUR SURVEY



- **A fourth random sample of 5: now the approval rating is lower than in population**
- **All of the polls were off by several percentage points**

OUR SURVEY



- **Let's keep doing this...**

OUR SURVEY

- Watch video at: https://www.dropbox.com/scl/fi/znyopmuehlr5i5za0jar6/psc202_f23.mp4

OUR SURVEY

- This was 500 random samples from the population
- Sample means (=different polls) jump around
- Some are far away from population mean, but most are quite close
- Mean of sample means gets pretty close to the population mean

AFTER 10,000 RANDOM SAMPLES