

PSC 202

SYRACUSE UNIVERSITY

INTRODUCTION TO POLITICAL ANALYSIS

MORE ON DESCRIBING VARIABLES
SAMPLING AND SURVEYS

SURVEY

- Take it if you haven't yet!
- Response rate: 80%
 - Need 85% to get extra credit for whole class
- rebrand.ly/202surveyfall23

EXAM

- **Next week Wednesday: Exam #1**
 - Can bring a calculator (no phone etc.)
 - Allowed to bring one single-page letter-size (8.5x11) sheet with you. Front side only. What you put on it is up to you, but it has to be your own.
- **Monday: Review**
 - Email questions etc. by Sunday evening
 - If you take exams at CDR, please sign up now!

EXAM

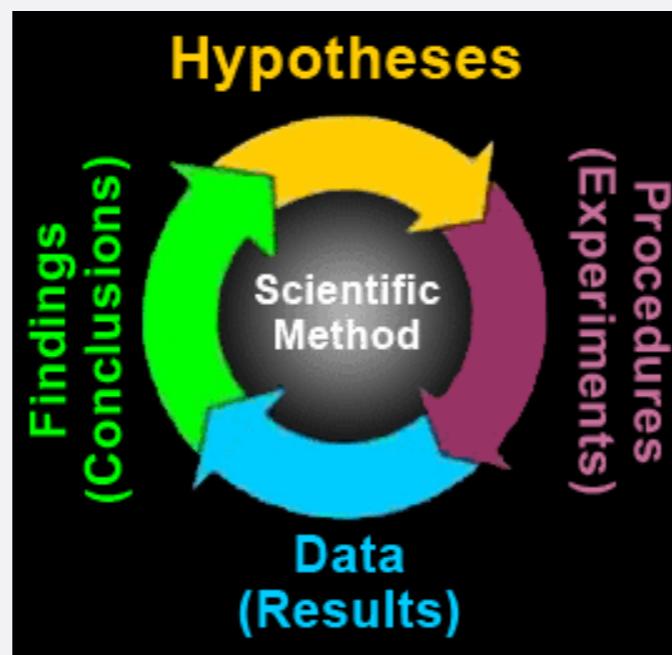
- **Reminder: Student Hours**
- **Monday 1-3 or by appointment**
 - Zoom or in person

PROBLEM SET 3

- Will post Problem Set 3 later
 - Due October 6 (Friday next week)
 - Good idea to attempt it before the exam (good practice)

WHERE WE ARE

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



LEVELS OF MEASUREMENT

	Nominal	Ordinal	Interval
Relative Differences	✓	✓	✓
Ranking	✗	✓	✓
Exact Differences	✗	✗	✓
Between Units			

CENTRAL TENDENCY

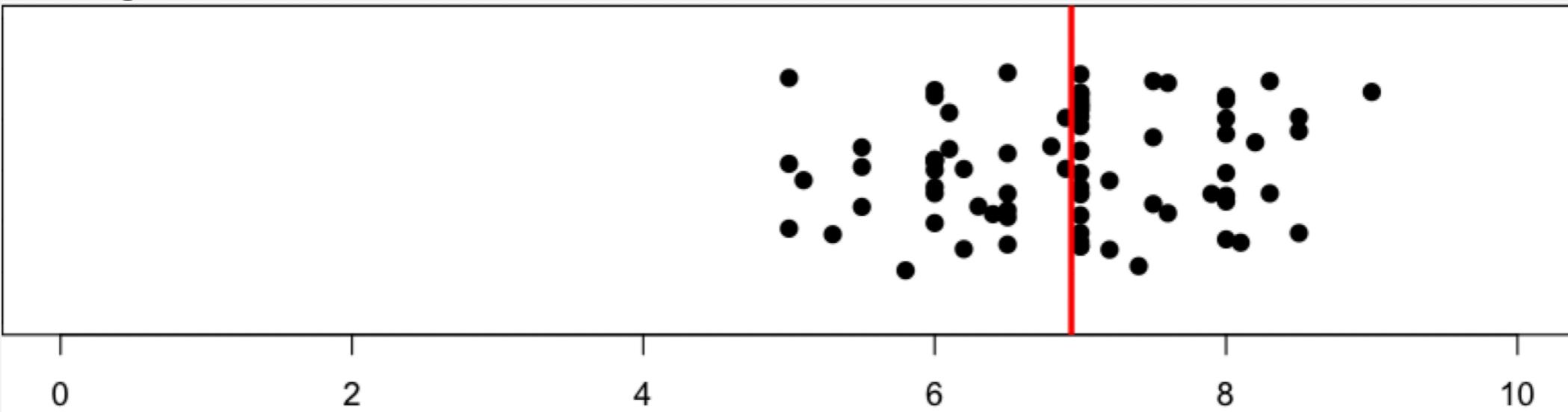
	Nominal	Ordinal	Interval
Mode	✓	✓	✓
Median	X	✓	✓
Mean	X	X	✓

FREQUENCY TABLE

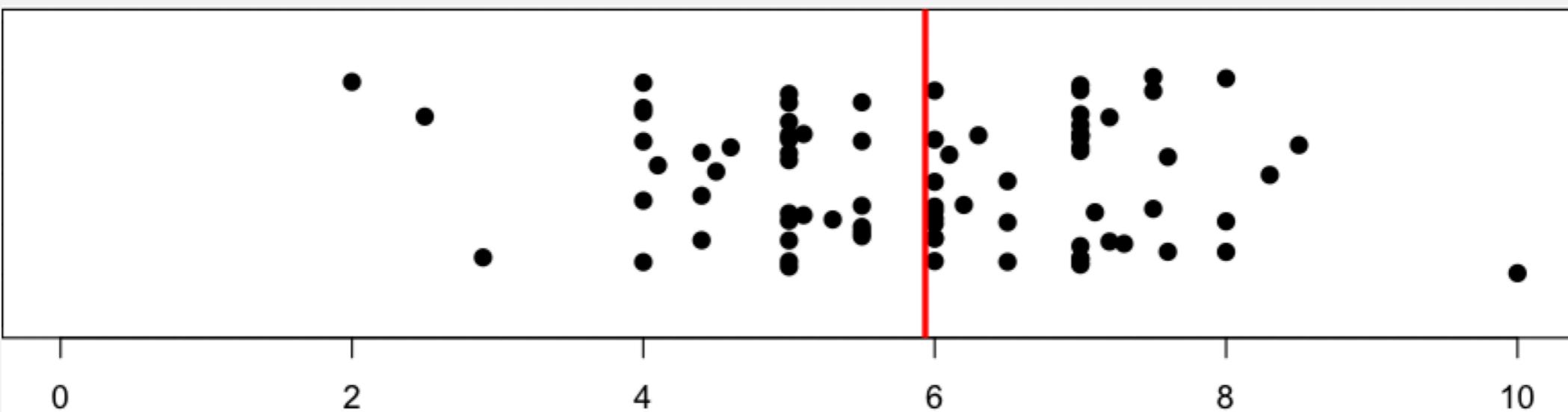
Age	Number	Percentage	Cumulative Percentage
18	7	8.5	8.5
19	51	62.2	70.7
20	13	15.6	86.3
21	8	9.8	96.1
22	1	1.3	97.4
24	1	1.3	98.7
27	1	1.3	100.0

SLEEP

- How many hours do you sleep at night?
 - Regular week:

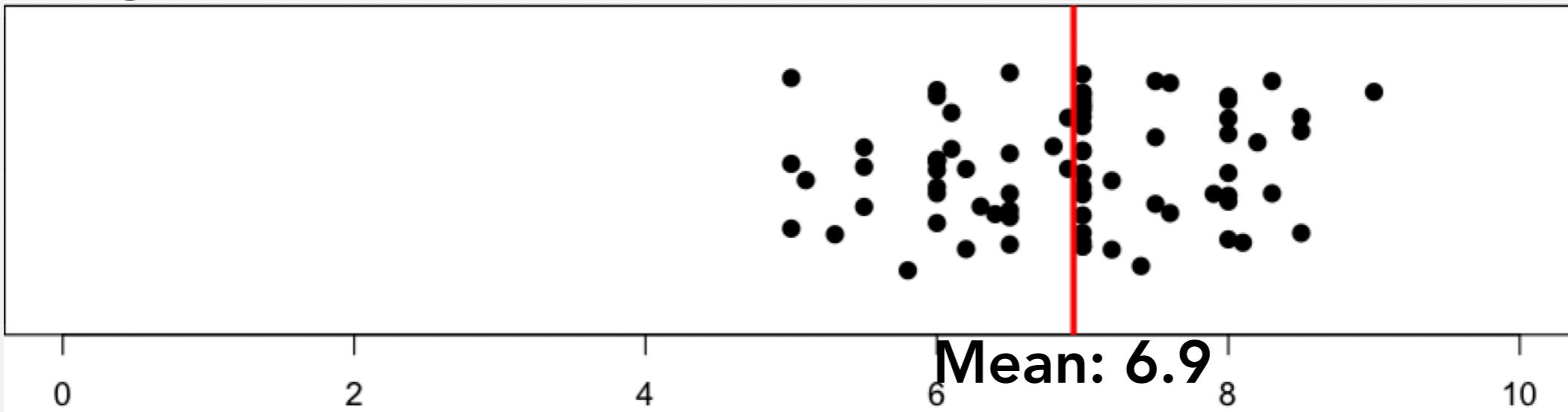


- Finals week:

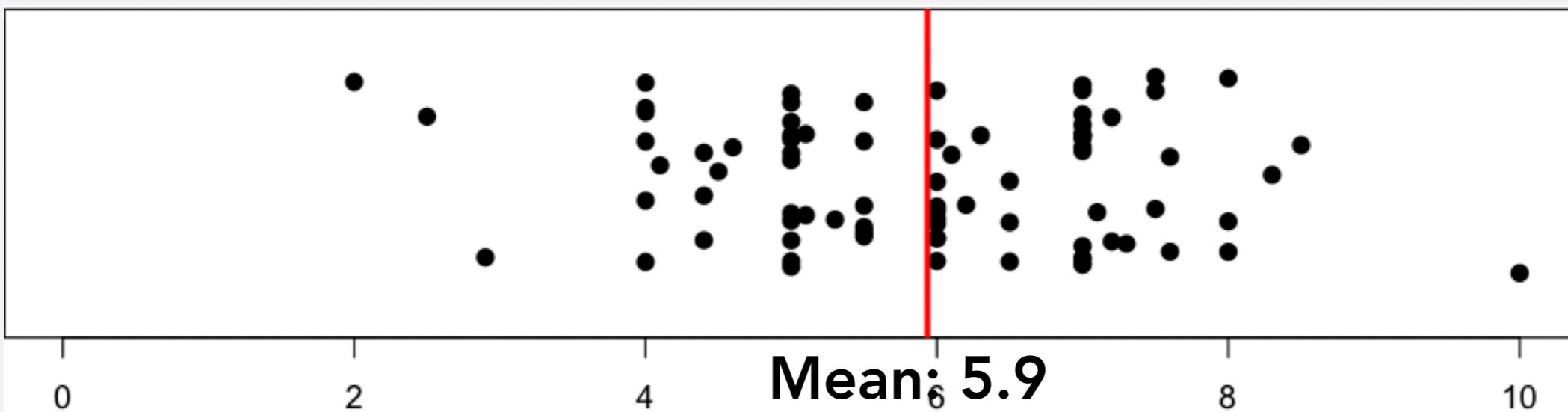


SLEEP

- How many hours do you sleep at night?
 - Regular week:

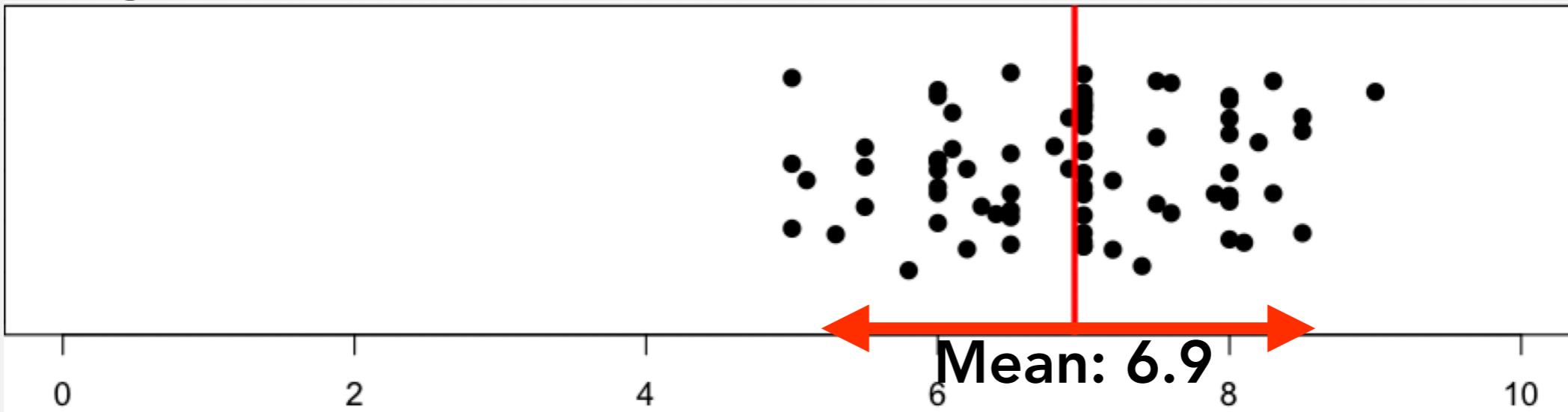


- Finals week:

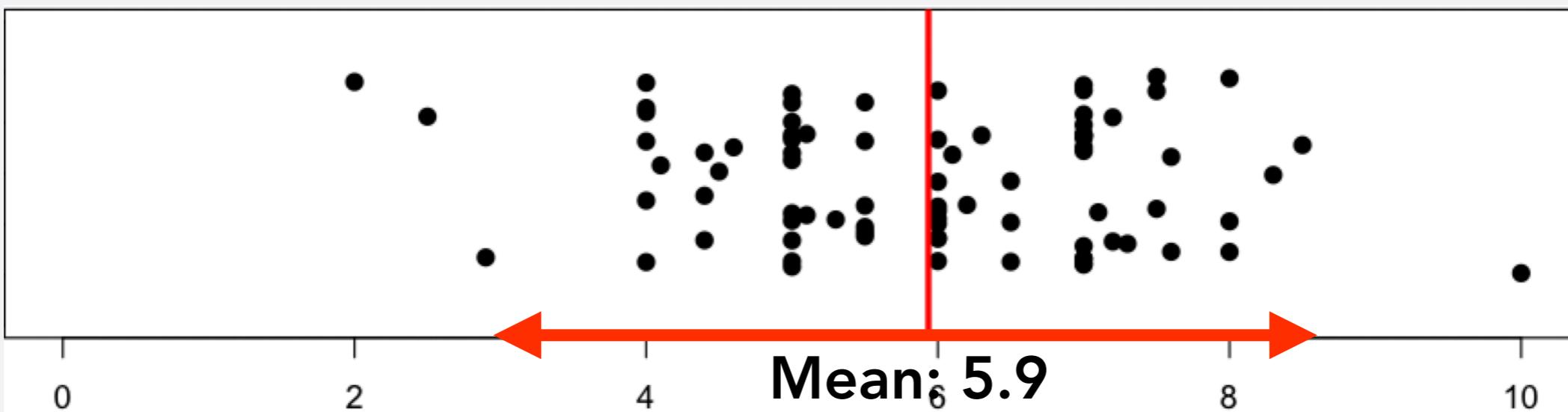


SLEEP

- How many hours do you sleep at night?
 - Regular week:

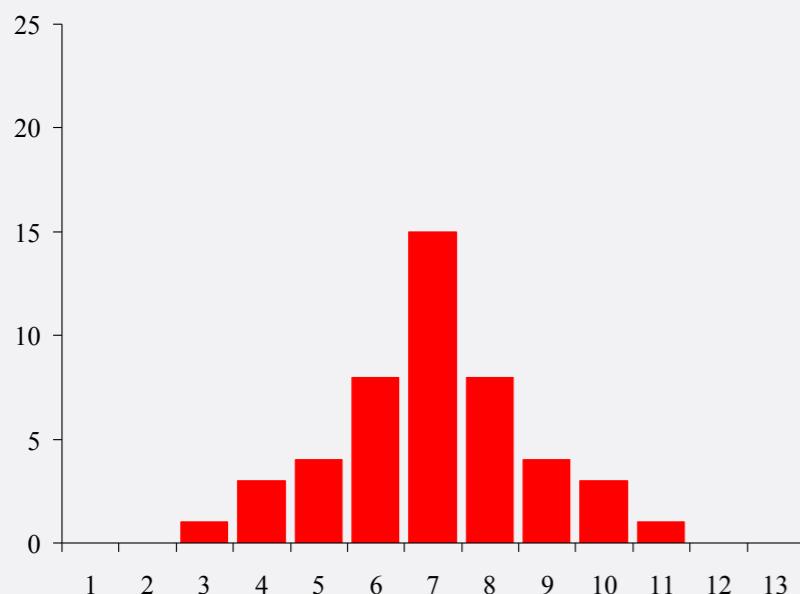


- Finals week:

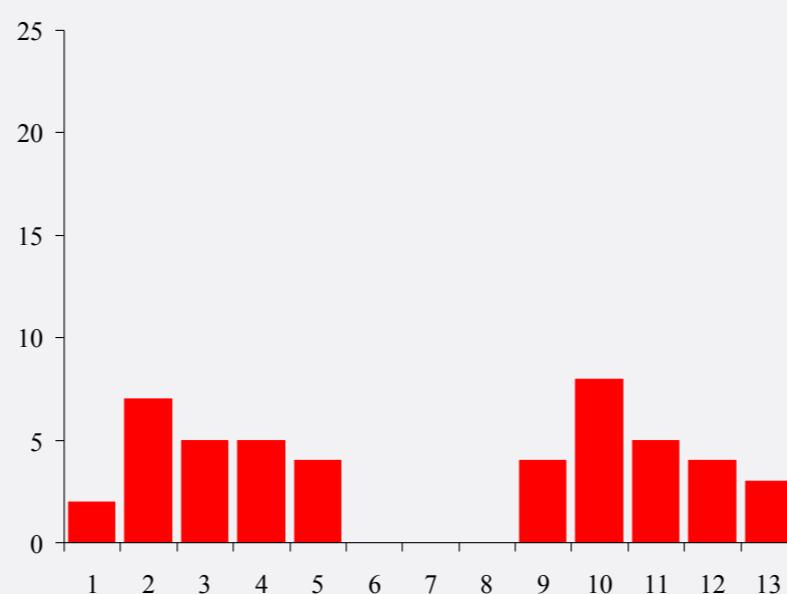


DISPERSION

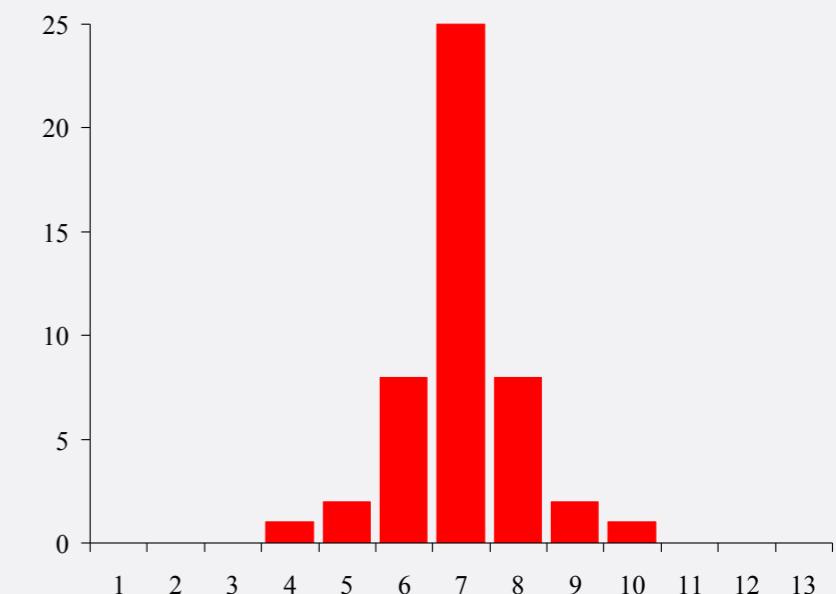
- All three: Median 7, Mean 7



Many students
perform mediocre,
some do well,
others not



One group does
very well, one group
does not



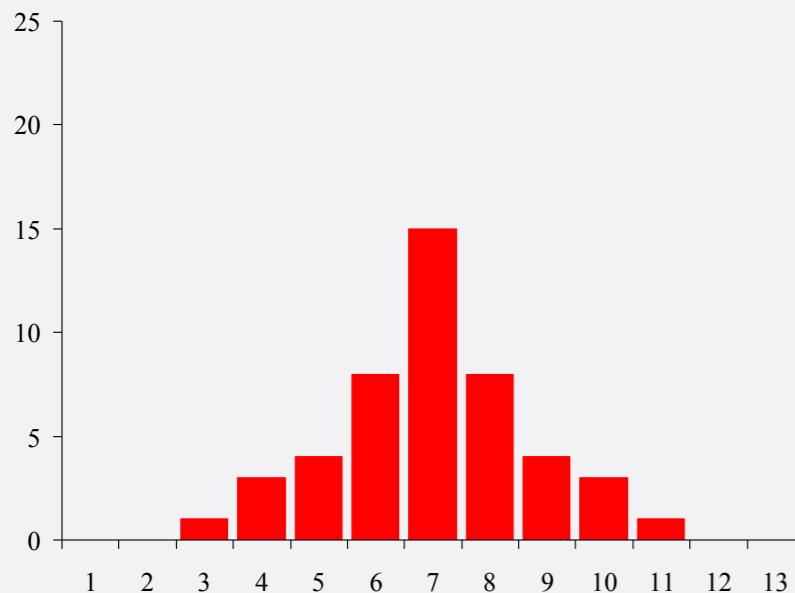
All students perform
relatively similarly
(mediocre)

DESCRIBING VARIABLES

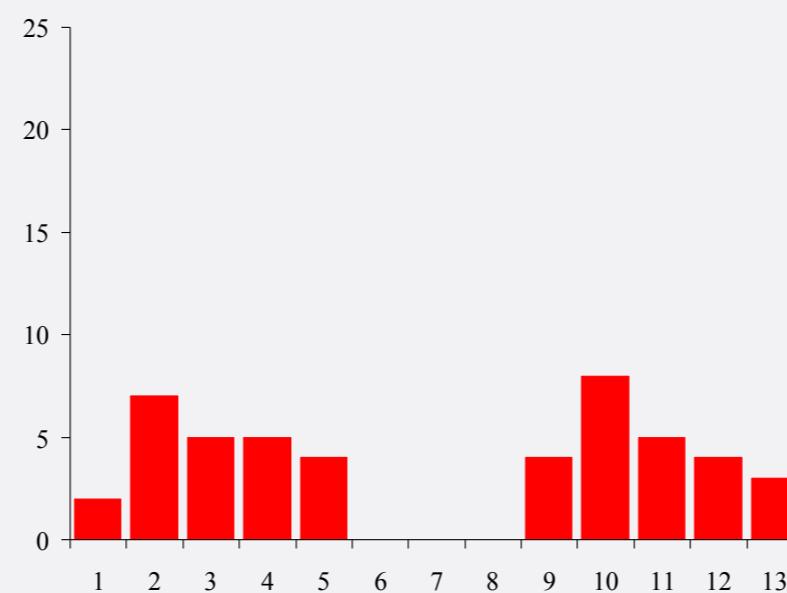
- Central tendency
- Frequency tables
- Dispersion

DISPERSION

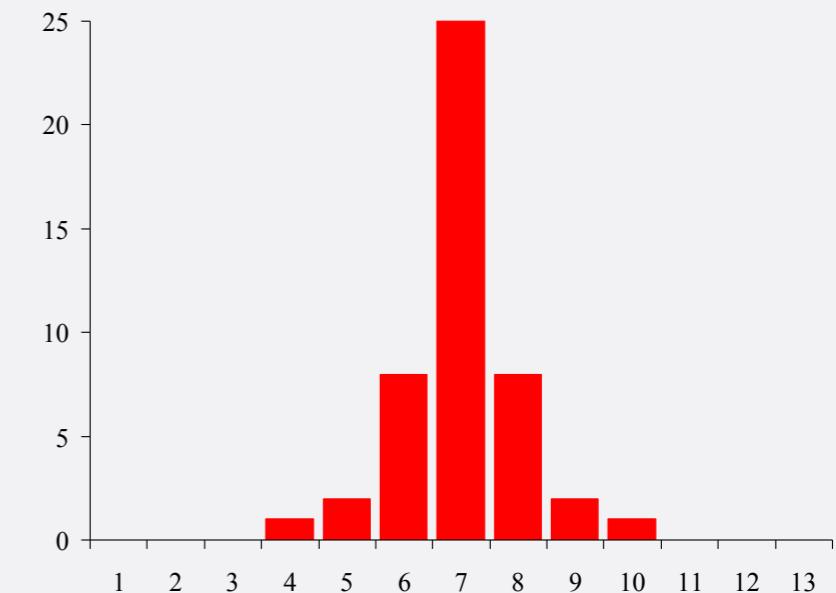
- Want to look at degree of variation around central tendency



Moderate variation



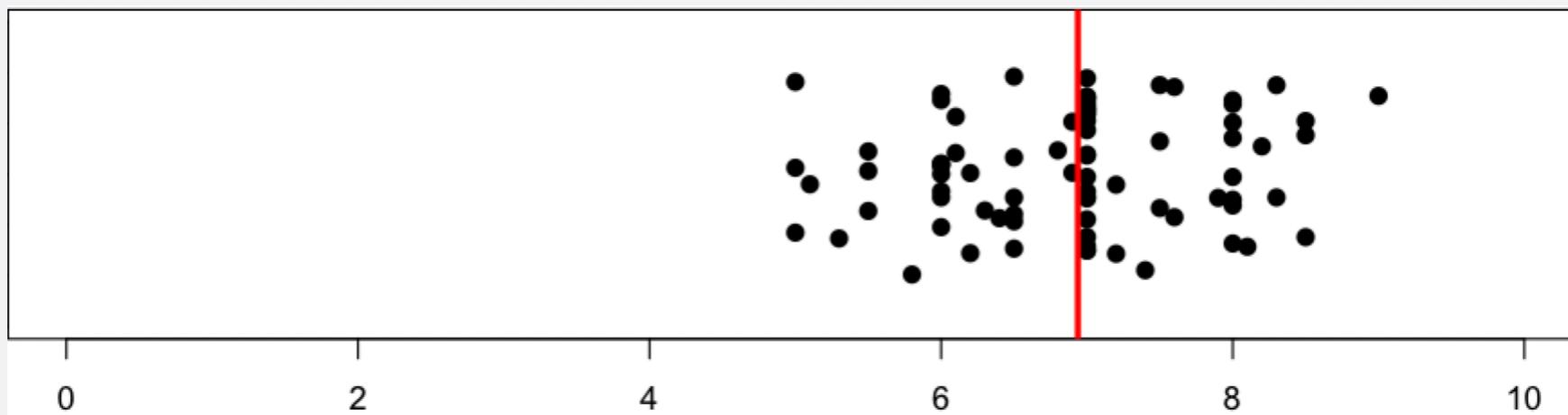
High variation



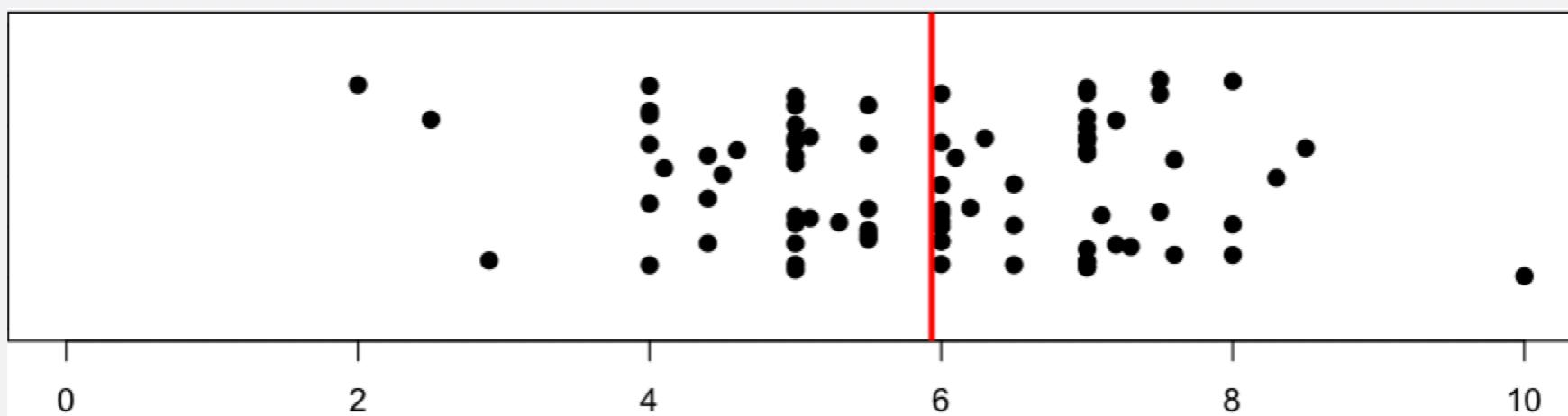
Low variation

DISPERSION

- Want to look at degree of variation around central tendency



Low variation



High variation

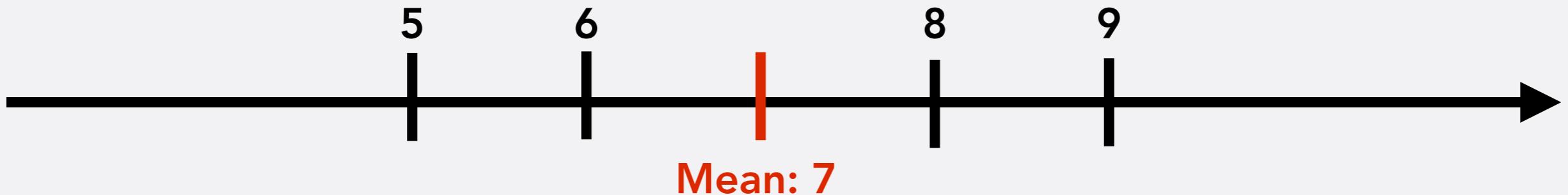
- How could we describe this mathematically?

DISPERSION

- Idea: Measure extent to which cases fall on or close to the mean of the distribution
- Easy measure of dispersion could be: Average distance of an observation from the mean

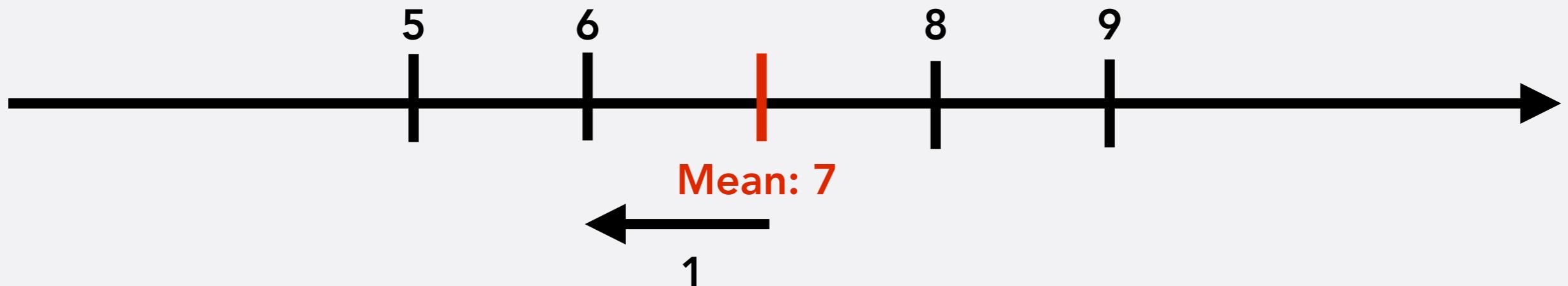
DISPERSION

- Simpler sleep example
 - 4 people: 5, 6, 8, 9 hours/night



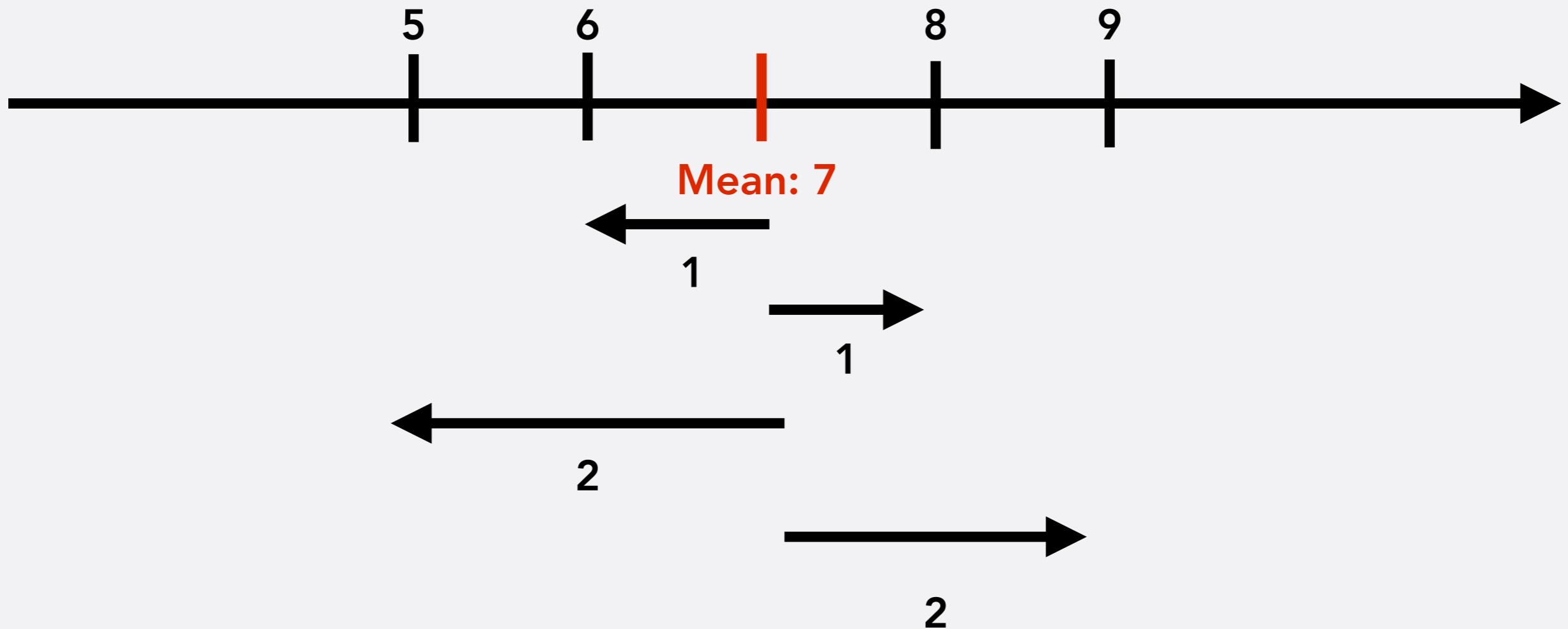
DISPERSION

- Simpler sleep example
 - 4 people: 5, 6, 8, 9 hours/night



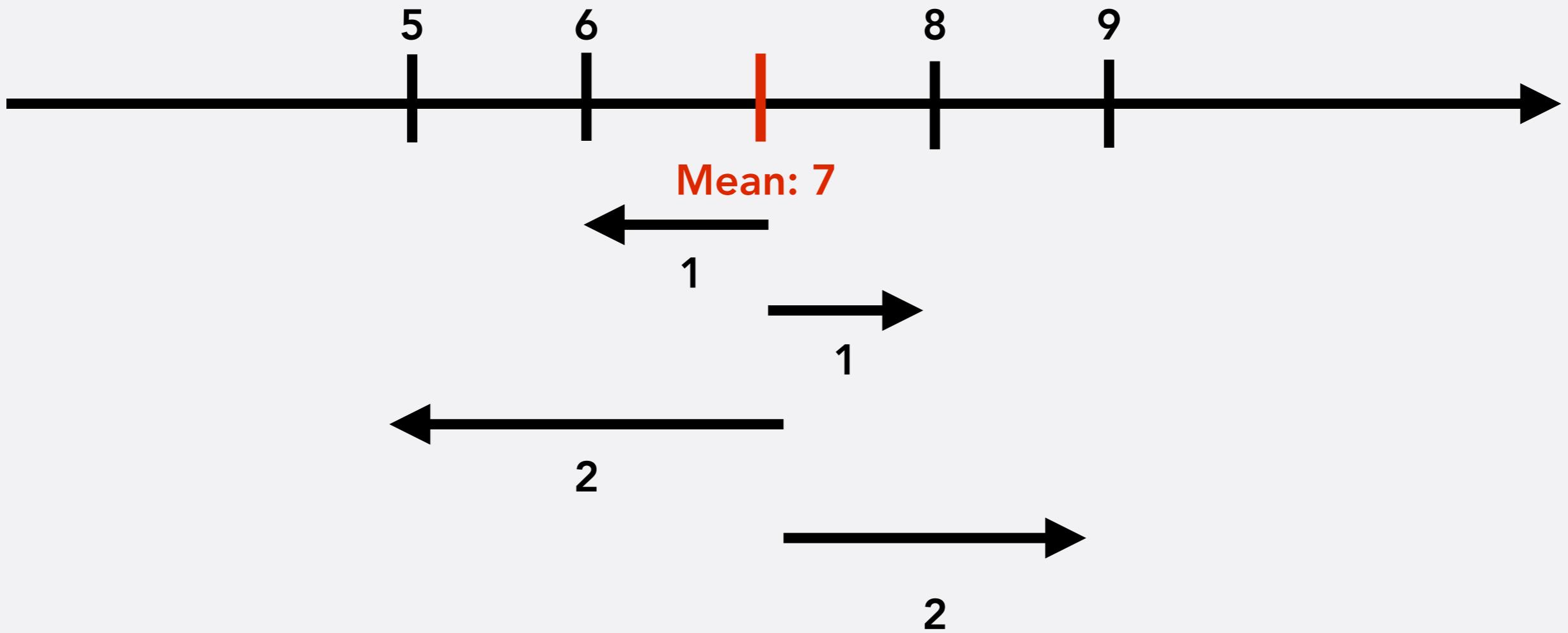
DISPERSION

- Simpler sleep example
 - 4 people: 5, 6, 8, 9 hours/night



DISPERSION

- Simpler sleep example
 - 4 people: 5, 6, 8, 9 hours/night



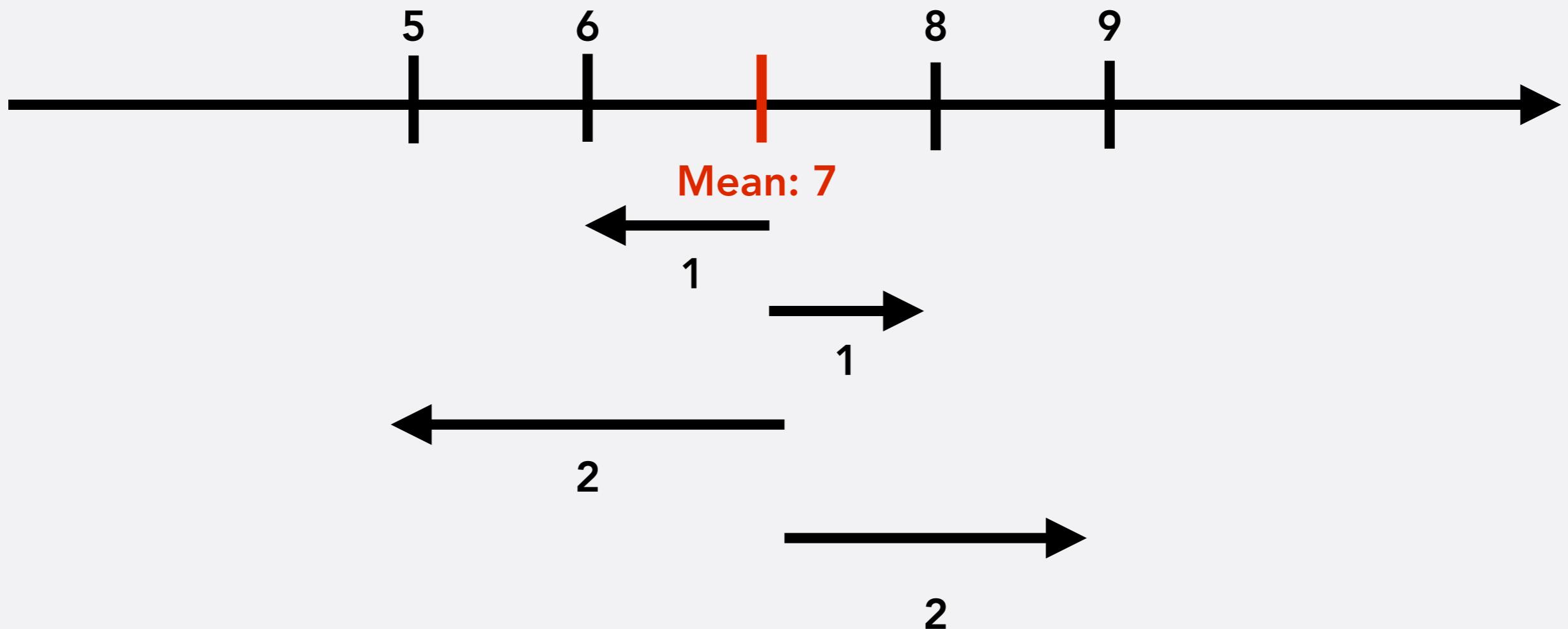
- Average distance from mean: $(1+1+2+2)/4=1.5$

STANDARD DEVIATION

- Unfortunately, this is *not* what we do
- Instead: standard deviation
 - Measures extent to which cases fall on or close to the mean of the distribution
 - *Kind of* measures the average distance of observations from the mean, but not quite
 - Gives extreme cases (far from mean) more weight

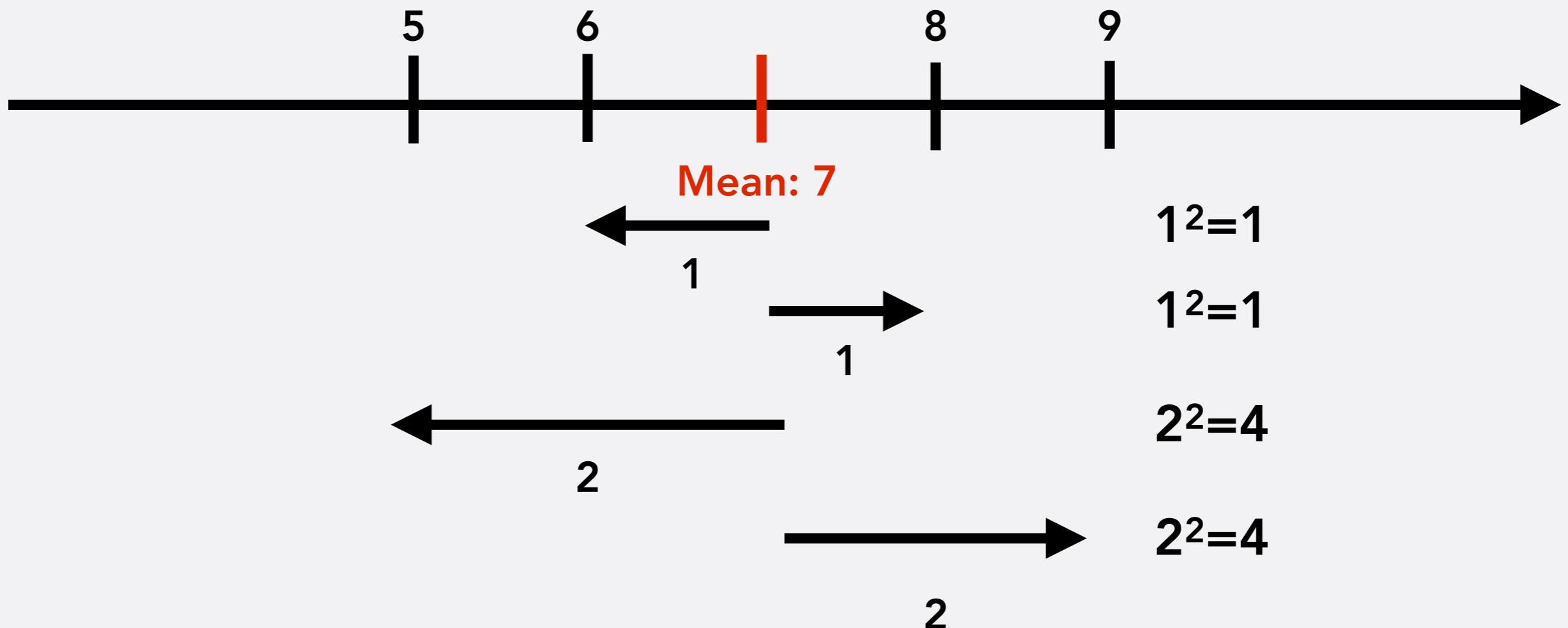
STANDARD DEVIATION

- We still start with taking the distance of each observation from the mean



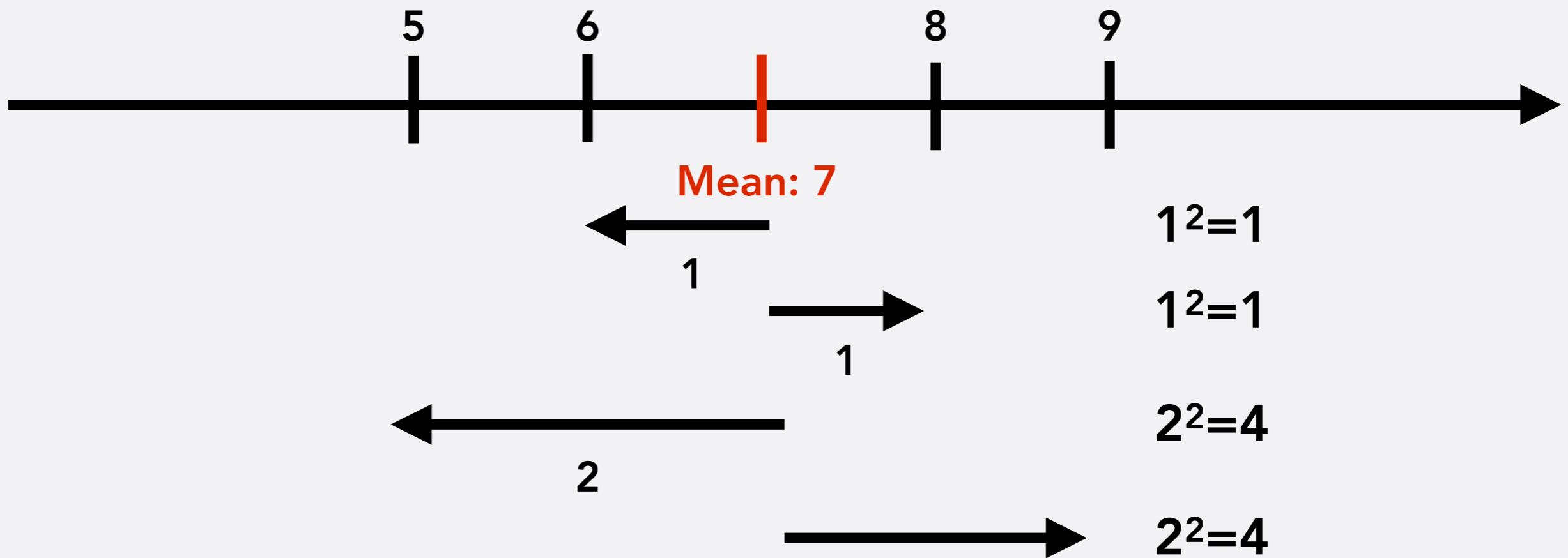
STANDARD DEVIATION

- But now we square each of them



STANDARD DEVIATION

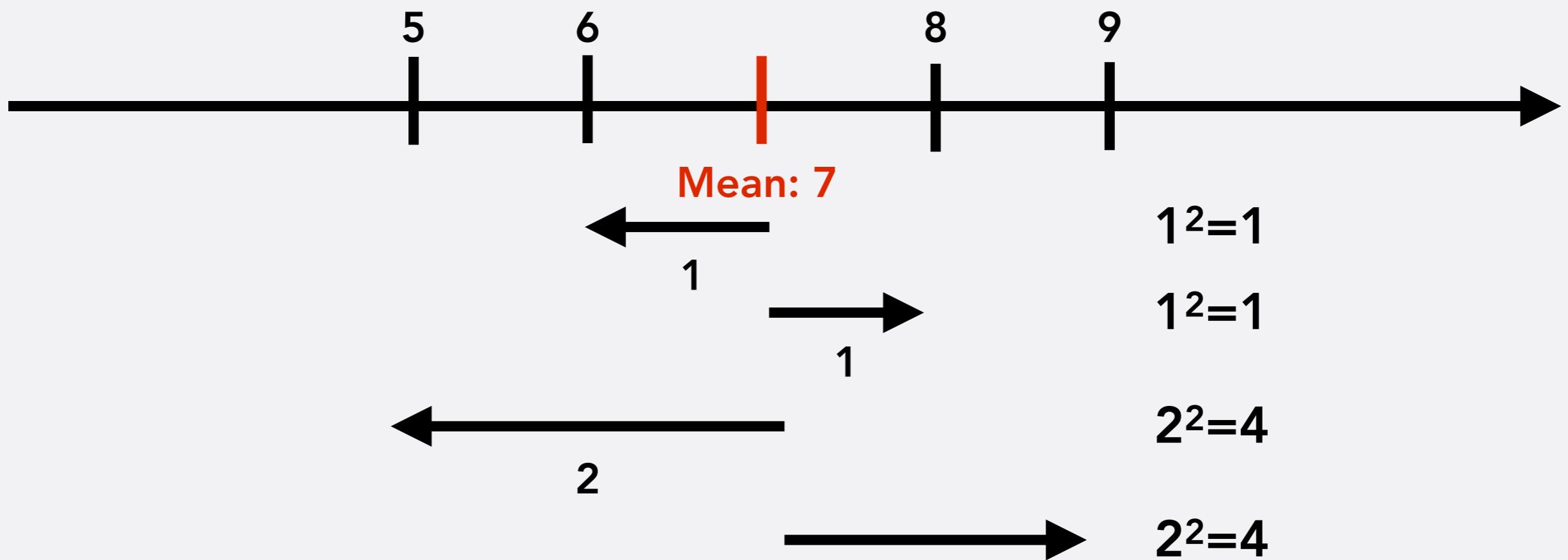
- Then take the average of those squared deviations



- $(1+1+4+4)/4=2.5$
- This is called the *variance*

STANDARD DEVIATION

- Finally, because we squared the deviations earlier, we now take the square root of the variance



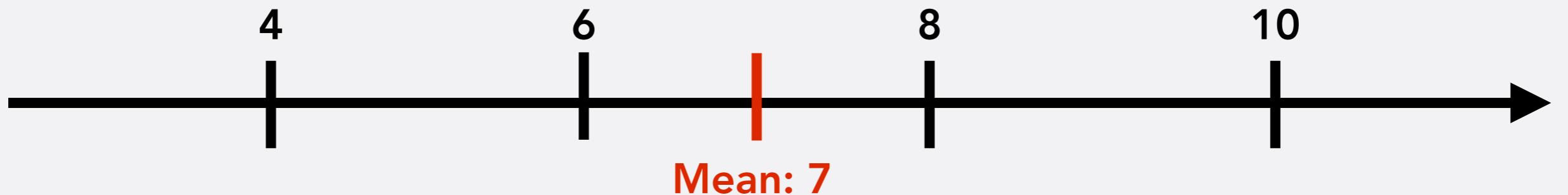
- $(1+1+4+4)/4=2.5$
- $\sqrt{2.5}=1.58$

STEPS

1. Calculate each value's deviation from mean
2. Square each deviation
3. Calculate the average of the sum of the squared deviations ("variance")
4. Take the square root of the variance ("standard deviation")

STANDARD DEVIATION

- Example with greater dispersion

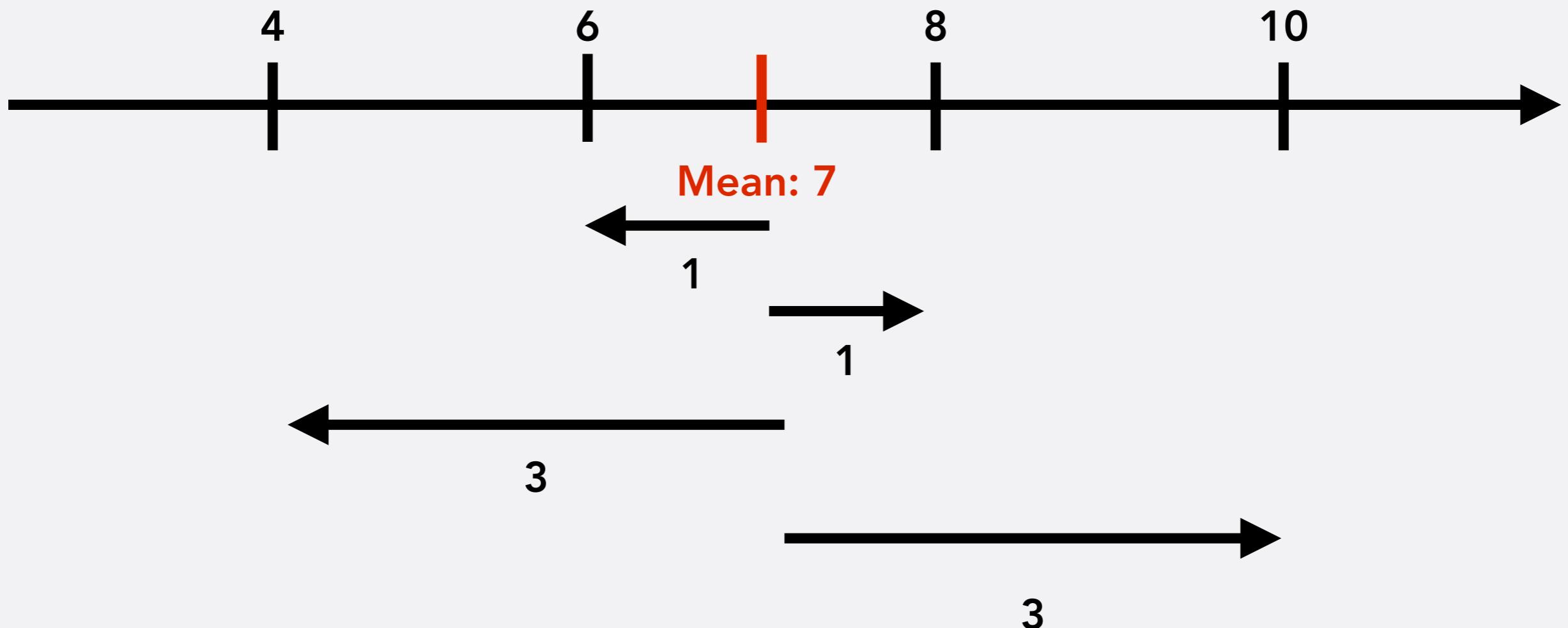


STEPS

1. Calculate each value's deviation from mean
2. Square each deviation
3. Calculate the average of the sum of the squared deviations ("variance")
4. Take the square root of the variance ("standard deviation")

STANDARD DEVIATION

- Example with greater dispersion

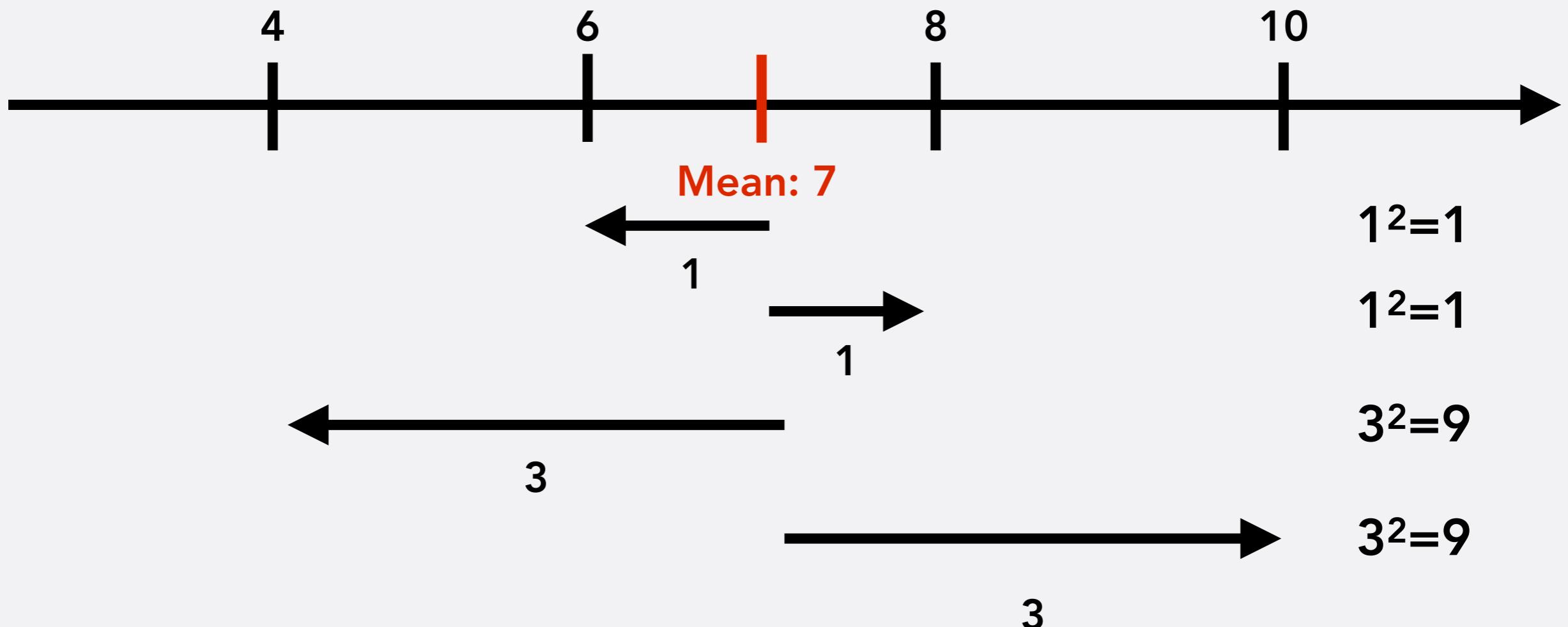


STEPS

1. Calculate each value's deviation from mean
2. Square each deviation
3. Calculate the average of the sum of the squared deviations ("variance")
4. Take the square root of the variance ("standard deviation")

STANDARD DEVIATION

- Example with greater dispersion

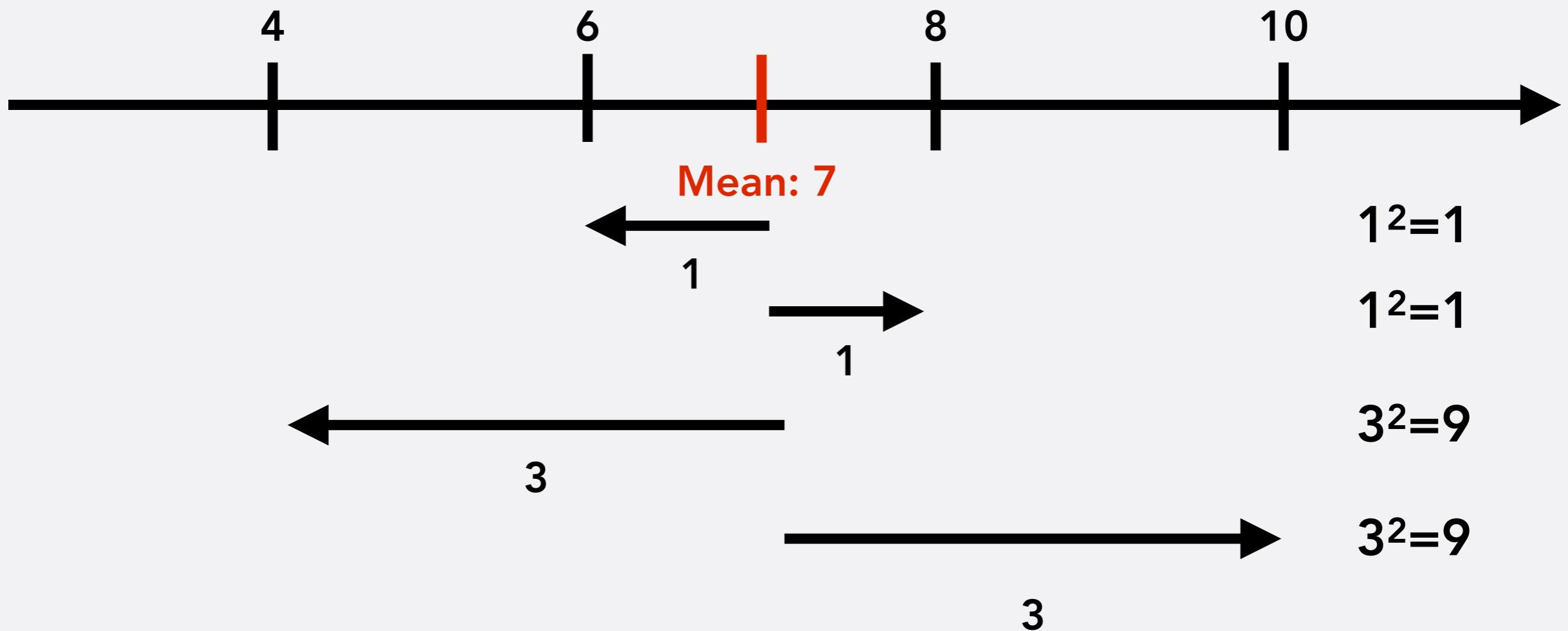


STEPS

1. Calculate each value's deviation from mean
2. Square each deviation
3. Calculate the average of the sum of the squared deviations ("variance")
4. Take the square root of the variance ("standard deviation")

STANDARD DEVIATION

- Example with greater dispersion



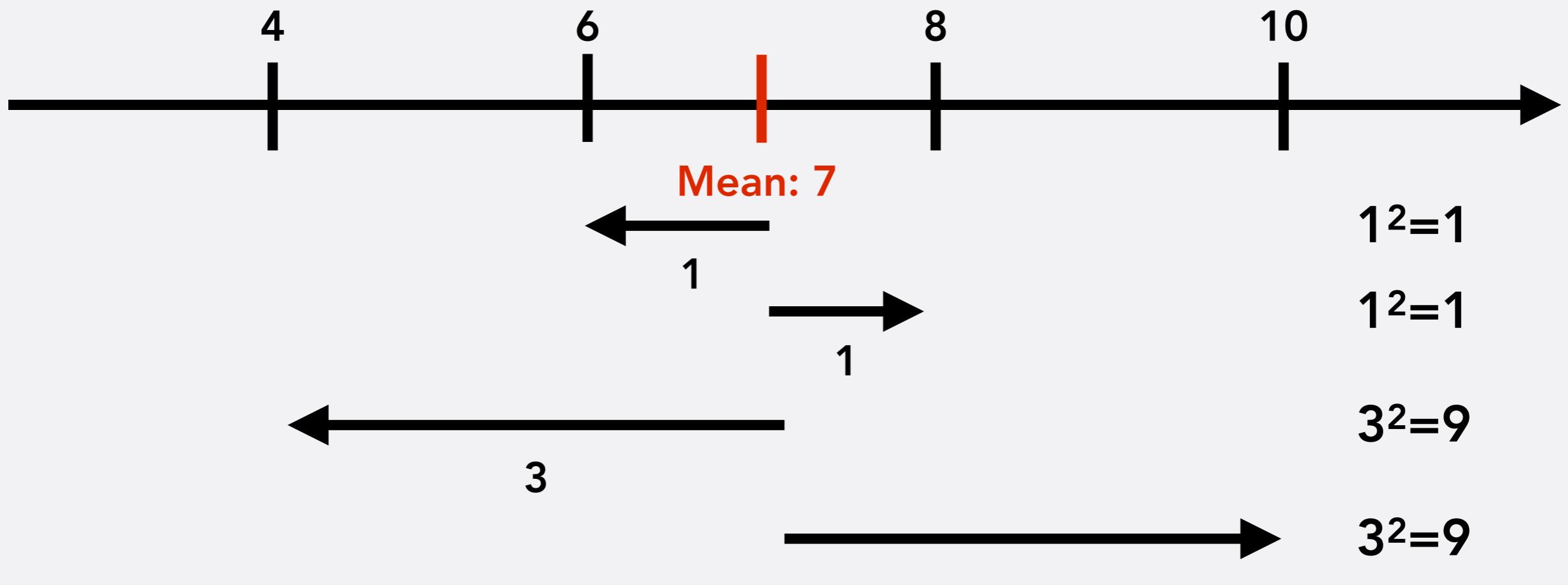
- $(1+1+9+9)/4=5$

STEPS

1. Calculate each value's deviation from mean
2. Square each deviation
3. Calculate the average of the sum of the squared deviations ("variance")
4. Take the square root of the variance ("standard deviation")

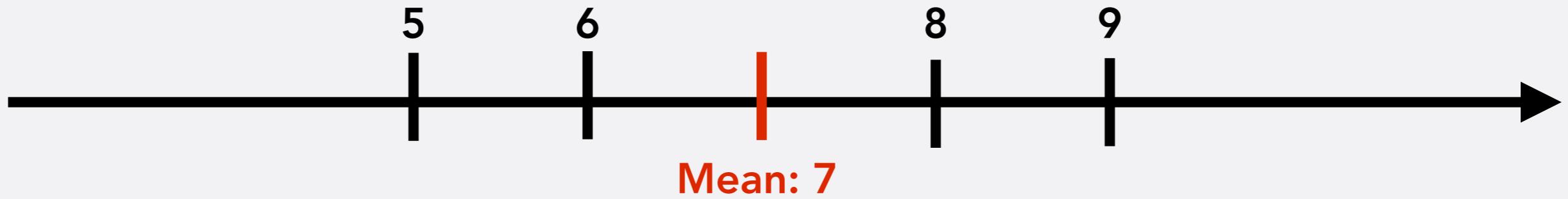
STANDARD DEVIATION

- Example with greater dispersion

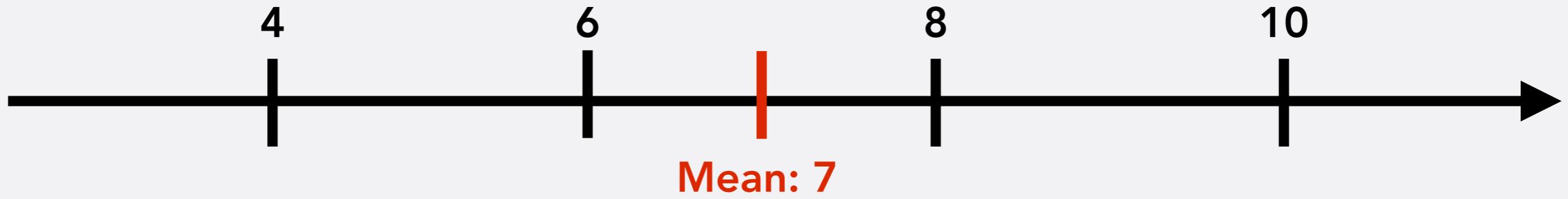


- $(1+1+9+9)/4=5$
- $\sqrt{5}=2.24$

COMPARISON



- Standard deviation: 1.58



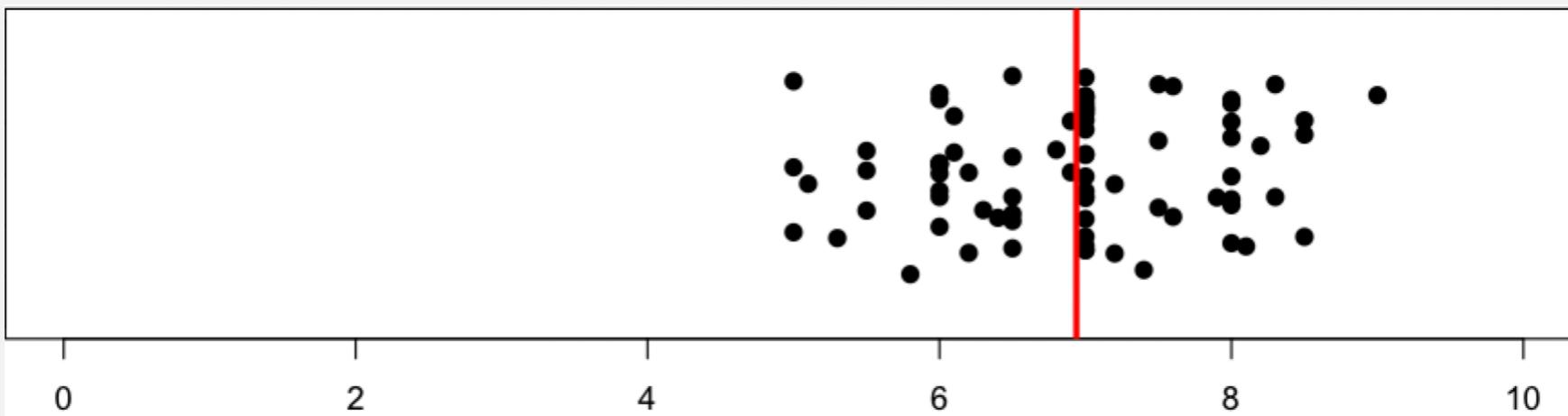
- Standard deviation: 2.24

STANDARD DEVIATION

- Standard deviation is helpful when comparing samples
 - Two countries with average income of 50k
 - One has a standard deviation of 5k, the other one a standard deviation of 15k

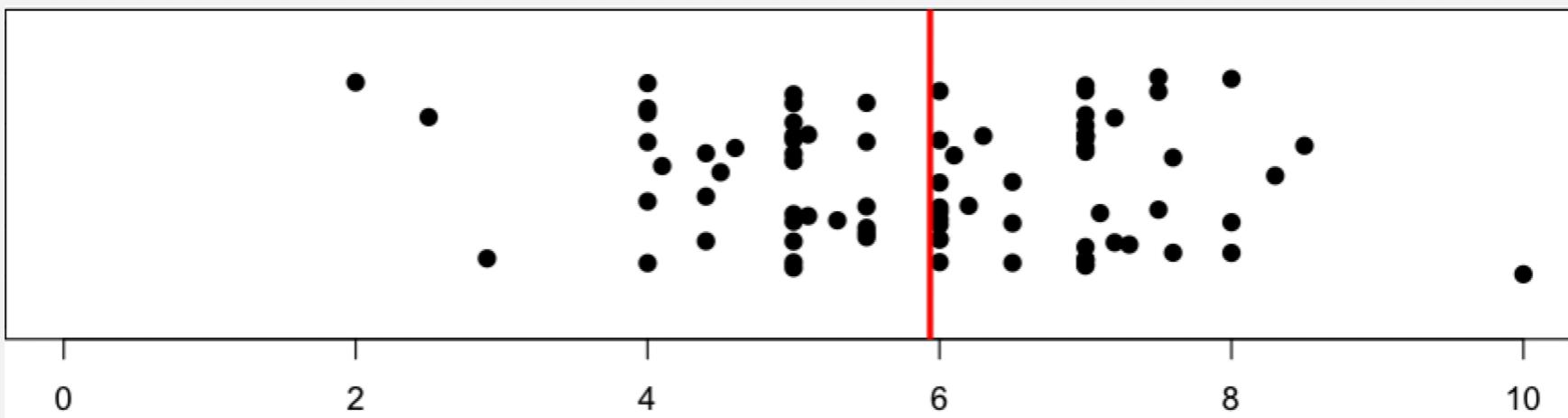
STANDARD DEVIATION

- How many hours do you sleep at night?
 - Regular week:



Mean: 6.9; Standard deviation: 1.00

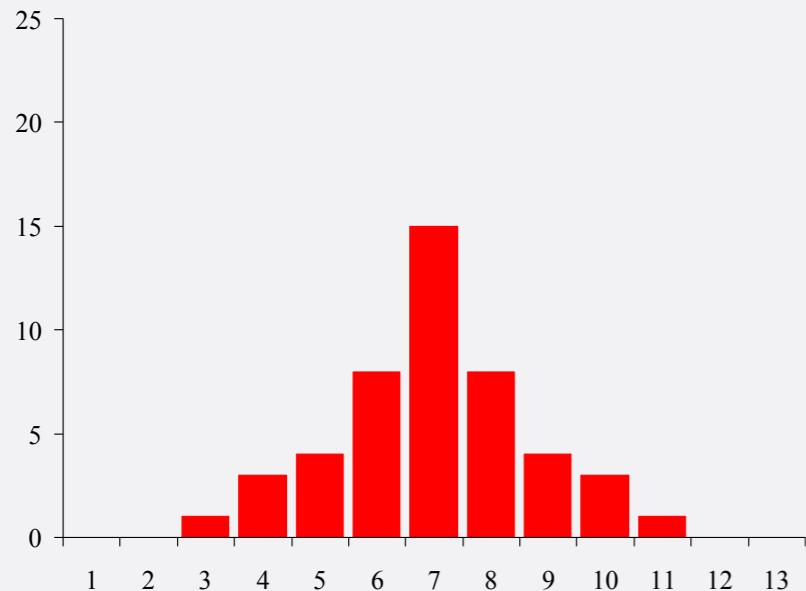
- Finals week:



Mean: 5.9; Standard deviation: 1.43

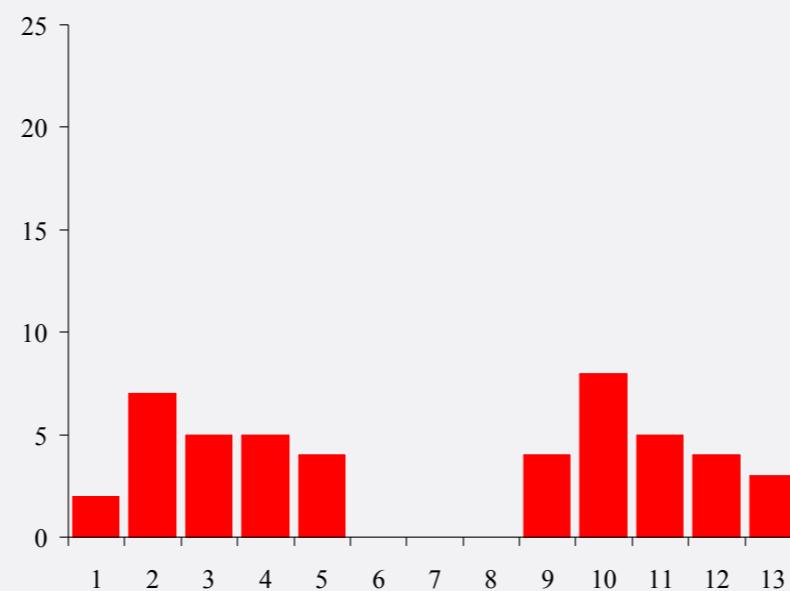
STANDARD DEVIATION

- Mean=7, median=7



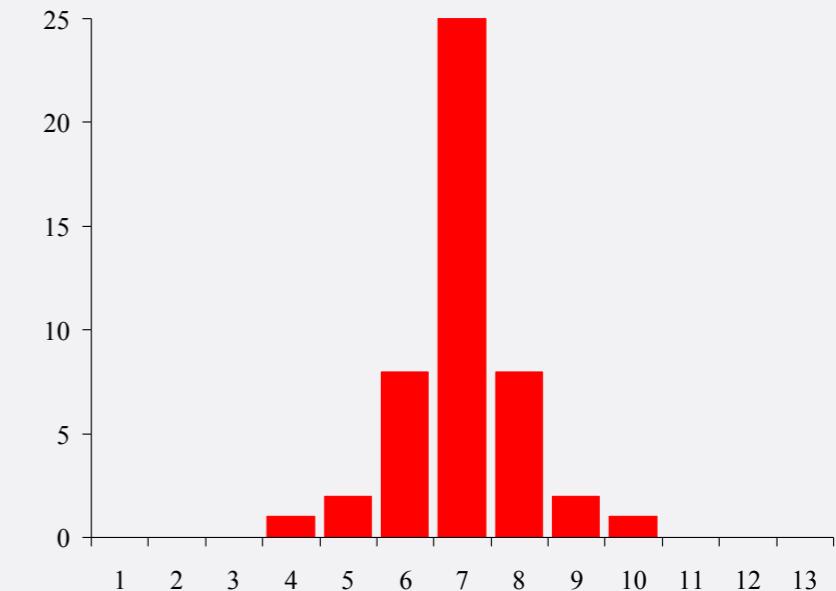
SD=1.67

Many students perform mediocre, some do well, others not



SD=4.01

One group does very well, one group does not



SD=1.02

All students perform relatively similarly (mediocre)

DISPERSION?

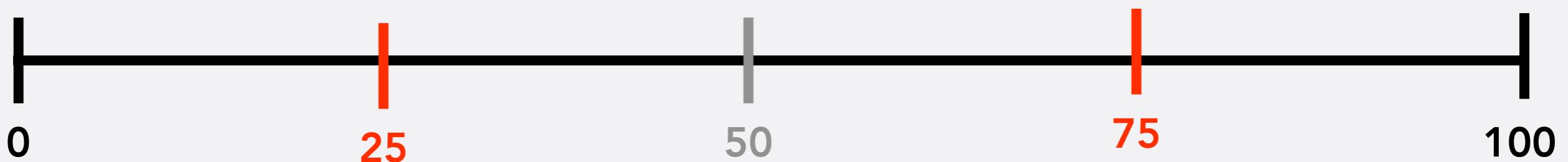
More Armed Security On Campus	Number	Percentage	Cumulative Percentage
Strongly agree	8	9.8	9.8
Somewhat agree	11	13.4	23.2
Neither agree nor disagree	20	24.4	47.6
Somewhat disagree	25	30.5	78.1
Strongly disagree	18	22.0	100.1

DISPERSION

- Standard deviation only works for interval-level variables
 - Here: Ordinal-level variable
- Dispersion measure: Interquartile range
 - Look at value at 25th percentile and value at 75th percentile

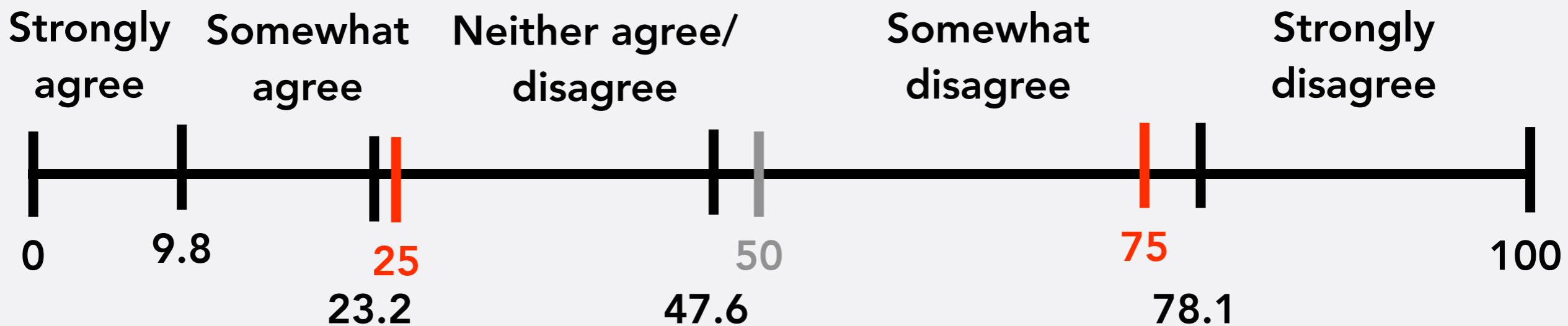
INTERQUARTILE RANGE

- More armed security on campus



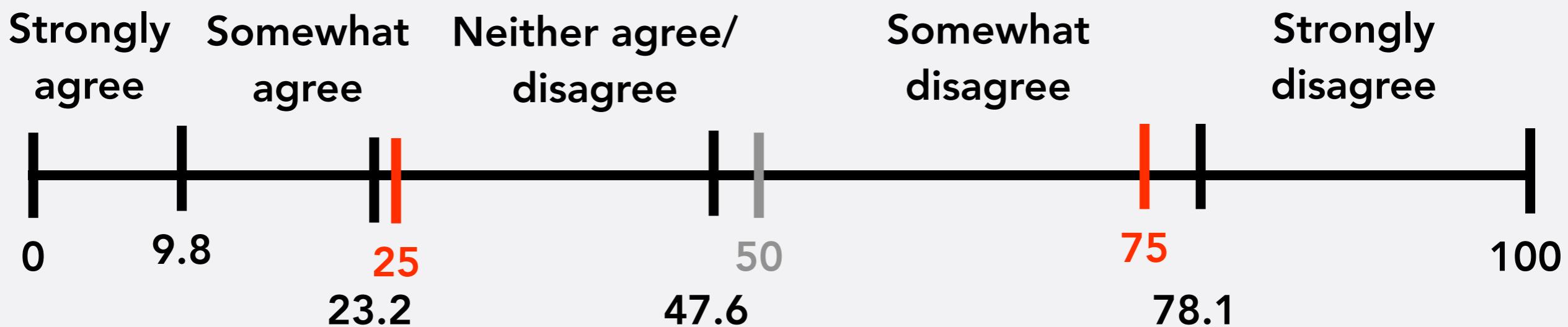
INTERQUARTILE RANGE

- More armed security on campus



INTERQUARTILE RANGE

- More armed security on campus



- 25th percentile: neither agree nor disagree
- 75th percentile: somewhat disagree

DISPERSION?

More Armed Security On Campus	Number	Percentage	Cumulative Percentage
Strongly agree	8	9.8	9.8
Somewhat agree	11	13.4	23.2
Neither agree nor disagree	20	24.4	47.6
Somewhat disagree	25	30.5	78.1
Strongly disagree	18	22.0	100.1

- Category whose cumulative percentage includes 25%?

DISPERSION?

25th

More Armed Security On Campus	Number	Percentage	Cumulative Percentage
Strongly agree	8	9.8	9.8
Somewhat agree	11	13.4	23.2
Neither agree nor disagree	20	24.4	47.6
Somewhat disagree	25	30.5	78.1
Strongly disagree	18	22.0	100.1

- Category whose cumulative percentage includes 25%?

DISPERSION?

25th

More Armed Security On Campus	Number	Percentage	Cumulative Percentage
Strongly agree	8	9.8	9.8
Somewhat agree	11	13.4	23.2
Neither agree nor disagree	20	24.4	47.6
Somewhat disagree	25	30.5	78.1
Strongly disagree	18	22.0	100.1

- Category whose cumulative percentage includes 75%?

DISPERSION?

More Armed Security On Campus	Number	Percentage	Cumulative Percentage
Strongly agree	8	9.8	9.8
Somewhat agree	11	13.4	23.2
Neither agree nor disagree	20	24.4	47.6
Somewhat disagree	25	30.5	78.1
Strongly disagree	18	22.0	100.1

- Category whose cumulative percentage includes 75%?

INTERQUARTILE RANGE

- **Interquartile range**
 - 25th percentile: “neither agree nor disagree”
 - 75th percentile: “somewhat disagree”
- **Works for ordinal and interval-level variables**
 - and can be interpreted in intuitive way

EXERCISE

Age	Number	Percentage	Cumulative Percentage
18	7	8.5	8.5
19	51	62.2	70.7
20	13	15.6	86.3
21	8	9.8	96.1
22	1	1.3	97.4
24	1	1.3	98.7
27	1	1.3	100.0

- Interquartile range? Standard deviation? Try at home!

NOMINAL-LEVEL

Party Affiliation	Number	Percentage
Democrat	53	64.6
Republican	2	2.4
Independent	14	17.1
Other	3	3.7
None	10	12.2

- No dispersion measure, since we can't rank observations

DESCRIBING VARIABLES

- There are different types of variables
 - Nominal, ordinal, interval
- Central tendency
 - Mode (n, o, i), median (o, i), mean (i)
- Dispersion
 - Standard deviation (i), interquartile range (o, i)

TODAY: SURVEYS AND SAMPLING

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.

WHAT THIS DOES AND DOESN'T TELL US

- **What this tells us:**
 - Biden's approval rating is 41% among 1,011 people who were interviewed

WHAT THIS DOES AND DOESN'T TELL US

- What we are *really* interested in
 - Approval rating for Biden among *all* American voters

TODAY AND NEXT CLASS

- How confident can we be that the 41% approval rating among 1,011 respondents is close to the approval rating among *all* American voters?

HOW IS THIS POSSIBLE?

Triumph of the Nerds: Nate Silver Wins in 50 States

 Share on Facebook

 Share on Twitter

+



Numbers nerd Nate Silver's forecasts prove all right on election night

FiveThirtyEight blogger predicted the outcome in all 50 states, assuming Barack Obama's Florida victory is confirmed

KEY DISTINCTION

- **Population:** the entire universe of objects to which our hypothesis applies (citizens, voters, countries, etc.)
 - Size of population: N
 - e.g. registered American voters, N=168,308,000

KEY DISTINCTION

- **Sample: the subset of the population that we study in order to make inferences about the full population**
 - Size of sample: n
 - e.g. respondents in the poll, n= 1,011

HOW TO SAMPLE

- Often we cannot study data from the entire population
 - Impractical, resource constraints
 - How do we get a sample from a population?

HOW TO SAMPLE

- 1936 Presidential Election: Franklin D. Roosevelt (D) vs. Alf Landon (R)
- Who will win?
- *Literary Digest* magazine poll
 - Sent 10 million mock ballots to people, addresses from subscriber data, phone records, automobile registrations
 - 2.5 million people returned ballot
- Prediction: Roosevelt 43%, Landon 57%

HOW TO SAMPLE

Straw Vote Fight Arouses Interest

Literary Digest and American Institute Are Far Apart

In Pre-election Forecast—Roosevelt, Landon

Both Get Around 56 Per Cent

- *American Institute for Public Opinion* poll
 - Interviewed 50,000 randomly sampled people
 - Prediction: Roosevelt 56%, Landon 44%

HOW TO SAMPLE

The Pittsburgh Press

WEATHER—Rain changing to snow and colder tonight. Thursday fair.

(Copyright, 1936, by Press Publishing Co. All Rights Reserved.)

VOLUME 53; No. 134 48 PAGES PITTSBURGH, PA., WEDNESDAY, NOVEMBER 4, 1936 DAILY NEWS-SENTINEL

**FINAL STOCKS
CLOSING PRICES
PRICE THREE CENTS**

ROOSEVELT WINS 46 STATES

*Landslide Gives Him Pennsylvania
By 600,000 And County By 188,000*



Nominee	Franklin D. Roosevelt	Alf Landon
Party	Democratic	Republican
Home state	New York	Kansas
Running mate	John Nance Garner	Frank Knox
Electoral vote	523	8
States carried	46	2
Popular vote	27,747,636	16,679,543
Percentage	60.8%	36.5%

WHAT WENT WRONG FOR LD?

WHAT WENT WRONG FOR LD?

- *Literary Digest* magazine poll
 - People who subscribe to magazines, own phones/cars are wealthy
 - Therefore more likely to vote for Republican
 - People who are less wealthy (and are more likely to for Democrat) were not in LD's sample
 - LD's sample was not representative of the population

WHAT WENT WRONG FOR LD?

- *American Institute* poll
 - Randomly selected people to interview
 - Its sample was representative of the population
 - It sampled poor and rich voters roughly in the same proportion as they are present in the population
 - Much more accurate prediction with 50,000 respondents than with 2.5 million of LD