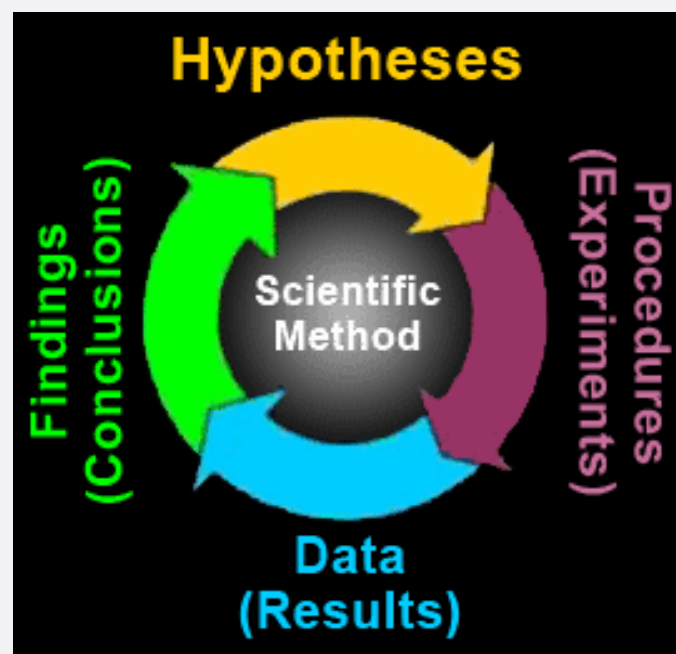# INTRODUCTION TO POLITICAL ANALYSIS

DESCRIBING VARIABLES

HOUSEKEEPING

- **Problem Set 3 will be posted tomorrow**
  - Due Friday next week (Feb 17)
- **Feb 20: Exam 1**

WHERE WE ARE

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

OPERATIONALIZATION

Concept

⬇

**Precise Definition of Concept
(Conceptual Definition)**

⬇

**Measurement Strategy
(Operational Definition)**

⬇

Variable

LEVELS OF MEASUREMENT

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

DESCRIBING VARIABLES

- **Central tendency**
- **Frequency tables**
- **Dispersion**

DESCRIBING A VARIABLE

	A	B	C	D
1	Respondent	Income	Marital Status	Universal Health Insurance
2	A	20,000	Single	Agree
3	B	55,000	Married	Disagree
4	C	35,000	Divorced	Strongly disagree
5	D	30,000	Married	Strongly disagree
6	E	80,000	Married	Strongly agree

- **How can we describe/summarize variables with many observations in a simple manner?**

DESCRIBING A VARIABLE

- What is a "typical" value?
- Central tendency

CENTRAL TENDENCY

	A	B	C	D
1	Respondent	Income	Marital Status	Universal Health Insurance
2	A	20,000	Single	Agree
3	B	55,000	Married	Disagree
4	C	35,000	Divorced	Strongly disagree
5	D	30,000	Married	Strongly disagree
6	E	80,000	Married	Strongly agree

- Income: What is the "typical income"?

CENTRAL TENDENCY

	A	B	C	D
1	Respondent	Income	Marital Status	Universal Health Insurance
2	A	20,000	Single	Agree
3	B	55,000	Married	Disagree
4	C	35,000	Divorced	Strongly disagree
5	D	30,000	Married	Strongly disagree
6	E	80,000	Married	Strongly agree

- Income: What is the "typical income"?
- Mean, average

CENTRAL TENDENCY

	A	B	C	D
1	Respondent	Income	Marital Status	Universal Health Insurance
2	A	20,000	Single	Agree
3	B	55,000	Married	Disagree
4	C	35,000	Divorced	Strongly disagree
5	D	30,000	Married	Strongly disagree
6	E	80,000	Married	Strongly agree

- Income: What is the "typical income"?
- Mean, average
- (20,000+55,000+35,000+30,000+80,000)/5=44,000

CENTRAL TENDENCY

	A	B	C	D
1	Respondent	Income	Marital Status	Universal Health Insurance
2	A	20,000	Single	Agree
3	B	55,000	Married	Disagree
4	C	35,000	Divorced	Strongly disagree
5	D	30,000	Married	Strongly disagree
6	E	80,000	Married	Strongly agree

- Universal Health Insurance: What is the "typical" value?

CENTRAL TENDENCY

	A	B	C	D
1	Respondent	Income	Marital Status	Universal Health Insurance
2	A	20,000	Single	Agree
3	B	55,000	Married	Disagree
4	C	35,000	Divorced	Strongly disagree
5	D	30,000	Married	Strongly disagree
6	E	80,000	Married	Strongly agree

- **Universal Health Insurance: What is the "typical" value?**
  - Cannot compute mean
  - How else could we describe the central tendency of this variable?

CENTRAL TENDENCY

	A	B	C	D
1	Respondent	Income	Marital Status	Universal Health Insurance
2	A	20,000	Single	Agree
3	B	55,000	Married	Disagree
4	C	35,000	Divorced	Strongly disagree
5	D	30,000	Married	Strongly disagree
6	E	80,000	Married	Strongly agree

- Line them up in order:
  - Strongly agree
  - Agree
  - Disagree
  - Strongly disagree
  - Strongly disagree

CENTRAL TENDENCY

	A	B	C	D
1	Respondent	Income	Marital Status	Universal Health Insurance
2	A	20,000	Single	Agree
3	B	55,000	Married	Disagree
4	C	35,000	Divorced	Strongly disagree
5	D	30,000	Married	Strongly disagree
6	E	80,000	Married	Strongly agree

- Line them up in order:
  - Strongly agree
  - Agree
  - Disagree
  - Strongly disagree
  - Strongly disagree
- "Median": Middle point of the variable

CENTRAL TENDENCY

	A	B	C	D
1	Respondent	Income	Marital Status	Universal Health Insurance
2	A	20,000	Single	Agree
3	B	55,000	Married	Disagree
4	C	35,000	Divorced	Strongly disagree
5	D	30,000	Married	Strongly disagree
6	E	80,000	Married	Strongly agree

- **What's the median of income?**

CENTRAL TENDENCY

	A	B	C	D
1	Respondent	Income	Marital Status	Universal Health Insurance
2	A	20,000	Single	Agree
3	B	55,000	Married	Disagree
4	C	35,000	Divorced	Strongly disagree
5	D	30,000	Married	Strongly disagree
6	E	80,000	Married	Strongly agree

- **What's the median of income?**
  - 20,000
  - 30,000
  - 35,000
  - 55,000
  - 80,000

CENTRAL TENDENCY

	A	B	C	D
1	Respondent	Income	Marital Status	Universal Health Insurance
2	A	20,000	Single	Agree
3	B	55,000	Married	Disagree
4	C	35,000	Divorced	Strongly disagree
5	D	30,000	Married	Strongly disagree
6	E	80,000	Married	Strongly agree

- What's the median of income?
  - 20,000
  - 30,000
  - 35,000
  - 55,000
  - 80,000

CENTRAL TENDENCY

	A	B	C	D
1	Respondent	Income	Marital Status	Universal Health Insurance
2	A	20,000	Single	Agree
3	B	55,000	Married	Disagree
4	C	35,000	Divorced	Strongly disagree
5	D	30,000	Married	Strongly disagree
6	E	80,000	Married	Strongly agree

- Marital Status: what is the "typical" marital status?

CENTRAL TENDENCY

	A	B	C	D
1	Respondent	Income	Marital Status	Universal Health Insurance
2	A	20,000	Single	Agree
3	B	55,000	Married	Disagree
4	C	35,000	Divorced	Strongly disagree
5	D	30,000	Married	Strongly disagree
6	E	80,000	Married	Strongly agree

- **Marital Status: what is the "typical" marital status?**
  - **Mean?**
  - **Median?**

CENTRAL TENDENCY

	A	B	C	D
1	Respondent	Income	Marital Status	Universal Health Insurance
2	A	20,000	Single	Agree
3	B	55,000	Married	Disagree
4	C	35,000	Divorced	Strongly disagree
5	D	30,000	Married	Strongly disagree
6	E	80,000	Married	Strongly agree

- **What is the most common value?**

CENTRAL TENDENCY

	A	B	C	D
1	Respondent	Income	Marital Status	Universal Health Insurance
2	A	20,000	Single	Agree
3	B	55,000	Married	Disagree
4	C	35,000	Divorced	Strongly disagree
5	D	30,000	Married	Strongly disagree
6	E	80,000	Married	Strongly agree

- **What is the most common value?**
  - Married
  - "Mode": most common value

CENTRAL TENDENCY

	A	B	C	D
1	Respondent	Income	Marital Status	Universal Health Insurance
2	A	20,000	Single	Agree
3	B	55,000	Married	Disagree
4	C	35,000	Divorced	Strongly disagree
5	D	30,000	Married	Strongly disagree
6	E	80,000	Married	Strongly agree

- What is the mode of marital status?
  -
- What is the mode of income?
  -

CENTRAL TENDENCY

	A	B	C	D
1	Respondent	Income	Marital Status	Universal Health Insurance
2	A	20,000	Single	Agree
3	B	55,000	Married	Disagree
4	C	35,000	Divorced	Strongly disagree
5	D	30,000	Married	Strongly disagree
6	E	80,000	Married	Strongly agree

- **What is the mode of marital status?**
  - Married
- **What is the mode of income?**
  - All of them are a mode

CENTRAL TENDENCY

- Can give mode for all variables
- For some variables, can provide median too
- For some, can even provide mean
- Why levels of measurement are important…

CENTRAL TENDENCY

	Nominal	Ordinal	Interval
Mode	√	√	√
Median	X	√	√
Mean	X	X	√

CENTRAL TENDENCY

	A	B	C	D
1	Respondent	Income	Marital Status	Universal Health Insurance
2	A	20,000	Single	Agree
3	B	55,000	Married	Disagree
4	C	35,000	Divorced	Strongly disagree
5	D	30,000	Married	Strongly disagree
6	E	80,000	Married	Strongly agree

- Income
  - Mode: all of them
  - Median: 35,000
  - Mean: 44,000
- Different central tendencies may give different results

CENTRAL TENDENCY

Respondent	Income	Marital Status	Universal Health Insurance
A	20,000	1	2
B	55,000	2	3
C	35,000	3	4
D	30,000	2	4
E	11,500,000,000	2	1

- Danger when using the mean: heavily influenced by outliers
- Suppose respondent E was Bill Gates

CENTRAL TENDENCY

Respondent	Income	Marital Status	Universal Health Insurance
A	20,000	1	2
B	55,000	2	3
C	35,000	3	4
D	30,000	2	4
E	11,500,000,000	2	1

- Danger when using the mean: heavily influenced by outliers
- Suppose respondent E was Bill Gates
- Mean: 2,300,028,000

CENTRAL TENDENCY

Respondent	Income	Marital Status	Universal Health Insurance
A	20,000	1	2
B	55,000	2	3
C	35,000	3	4
D	30,000	2	4
E	11,500,000,000	2	1

- Danger when using the mean: heavily influenced by outliers
- Suppose respondent E was Bill Gates
- Mean: 2,300,028,000
- Good reflection of "central tendency"?

CENTRAL TENDENCY

Respondent	Income	Marital Status	Universal Health Insurance
A	20,000	1	2
B	55,000	2	3
C	35,000	3	4
D	30,000	2	4
E	11,500,000,000	2	1

- What's the median income?

CENTRAL TENDENCY

Respondent	Income	Marital Status	Universal Health Insurance
A	20,000	1	2
B	55,000	2	3
C	35,000	3	4
D	30,000	2	4
E	11,500,000,000	2	1

- What's the median income?
  - 20,000
  - 30,000
  - 35,000
  - 55,000
  - 11,500,000,000
- In this case, describes central tendency better

CENTRAL TENDENCY

	Mode	Median	Mean
Nominal	√	X	X
Ordinal	√	√	X
Interval	√	√	√

DESCRIBING VARIABLES

- **Central tendency**
- **Frequency tables**
- **Dispersion**

FREQUENCY TABLE

Primary News Source	Number	Percentage
Internet (no social media)	66	56.4
Social Media	24	20.5
Print	13	11.1
TV	6	5.1
Radio	4	3.4
Conversations	2	1.7
Other	2	1.7

- **Nominal-level variable**

FREQUENCY TABLE

Cannabis Should Be Legalized	Number	Percentage	Cumulative Percentage
Strongly agree	71	60.7	60.7
Somewhat agree	24	20.5	81.2
Neither agree nor disagree	10	8.5	89.7
Somewhat disagree	5	4.3	94.0
Strongly disagree	7	6.0	100.0

- **Ordinal-level variable**

FREQUENCY TABLE

Age	Number	Percentage	Cumulative Percentage
17	1	0.9	0.9
18	27	23.3	24.2
19	42	36.2	60.4
20	25	21.6	82.0
21	17	14.7	96.7
22	4	3.4	100.1

- **Interval-level variable**

CENTRAL TENDENCY

- So far: computing central tendencies from list of observations
  - 4, 4, 5, 6, 7
  - mean: (4+4+5+6+7)/5=5.2
  - median: 5
  - mode: 4
- now: compute central tendencies from frequency table

FREQUENCY TABLE

Primary News Source	Number	Percentage
Internet (no social media)	66	56.4
Social Media	24	20.5
Print	13	11.1
TV	6	5.1
Radio	4	3.4
Conversations	2	1.7
Other	2	1.7

- **What central tendencies can we compute?**

FREQUENCY TABLE

Cannabis Should Be Legalized	Number	Percentage	Cumulative Percentage
Strongly agree	71	60.7	60.7
Somewhat agree	24	20.5	81.2
Neither agree nor disagree	10	8.5	89.7
Somewhat disagree	5	4.3	94.0
Strongly disagree	7	6.0	100.0

- **What central tendencies can we compute?**

MEDIAN

- **Cannabis should be legalized**

0   50   100

FREQUENCY TABLE

Cannabis Should Be Legalized	Number	Percentage	Cumulative Percentage
Strongly agree	71	60.7	60.7
Somewhat agree	24	20.5	81.2
Neither agree nor disagree	10	8.5	89.7
Somewhat disagree	5	4.3	94.0
Strongly disagree	7	6.0	100.0

- **What central tendencies can we compute?**

MEDIAN

- **Cannabis should be legalized**

Strongly
agree

0   50   60.7   100

FREQUENCY TABLE

Cannabis Should Be Legalized	Number	Percentage	Cumulative Percentage
Strongly agree	71	60.7	60.7
Somewhat agree	24	20.5	81.2
Neither agree nor disagree	10	8.5	89.7
Somewhat disagree	5	4.3	94.0
Strongly disagree	7	6.0	100.0

- **What central tendencies can we compute?**

MEDIAN

- **Cannabis should be legalized**

FREQUENCY TABLE

Cannabis Should Be Legalized	Number	Percentage	Cumulative Percentage
Strongly agree	71	60.7	60.7
Somewhat agree	24	20.5	81.2
Neither agree nor disagree	10	8.5	89.7
Somewhat disagree	5	4.3	94.0
Strongly disagree	7	6.0	100.0

- **What central tendencies can we compute?**

MEDIAN

- **Cannabis should be legalized**

FREQUENCY TABLE

Cannabis Should Be Legalized	Number	Percentage	Cumulative Percentage
Strongly agree	71	60.7	60.7
Somewhat agree	24	20.5	81.2
Neither agree nor disagree	10	8.5	89.7
Somewhat disagree	5	4.3	94.0
Strongly disagree	7	6.0	100.0

- **What central tendencies can we compute?**

MEDIAN

- **Cannabis should be legalized**
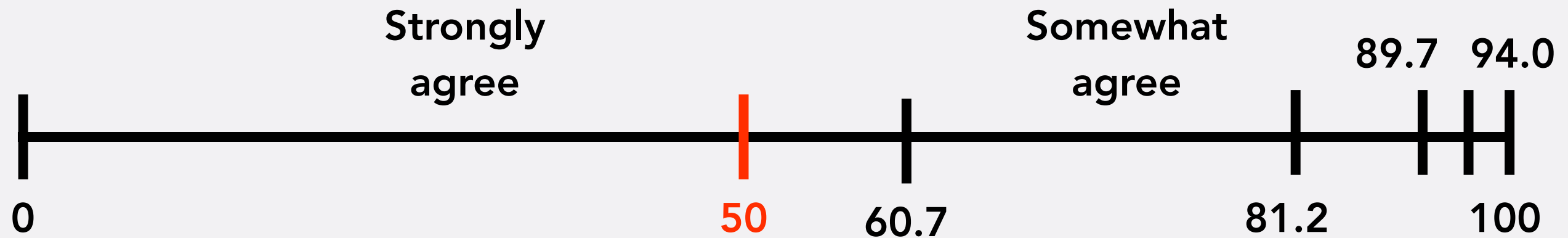
- **Median: strongly agree**

FREQUENCY TABLE

Cannabis Should Be Legalized	Number	Percentage	Cumulative Percentage
Strongly agree	71	60.7	60.7
Somewhat agree	24	20.5	81.2
Neither agree nor disagree	10	8.5	89.7
Somewhat disagree	5	4.3	94.0
Strongly disagree	7	6.0	100.0

- **Median: Category whose cumulative percentage includes 50%**

FREQUENCY TABLE

Age	Number	Percentage	Cumulative Percentage
17	1	0.9	0.9
18	27	23.3	24.2
19	42	36.2	60.4
20	25	21.6	82.0
21	17	14.7	96.7
22	4	3.4	100.1

- **What central tendencies can we compute?**

MEDIAN

- Age

18yrs   19yrs   20yrs   21yrs 22yrs

0   24.2   50   60.4   82.0   100
                                                    96.7
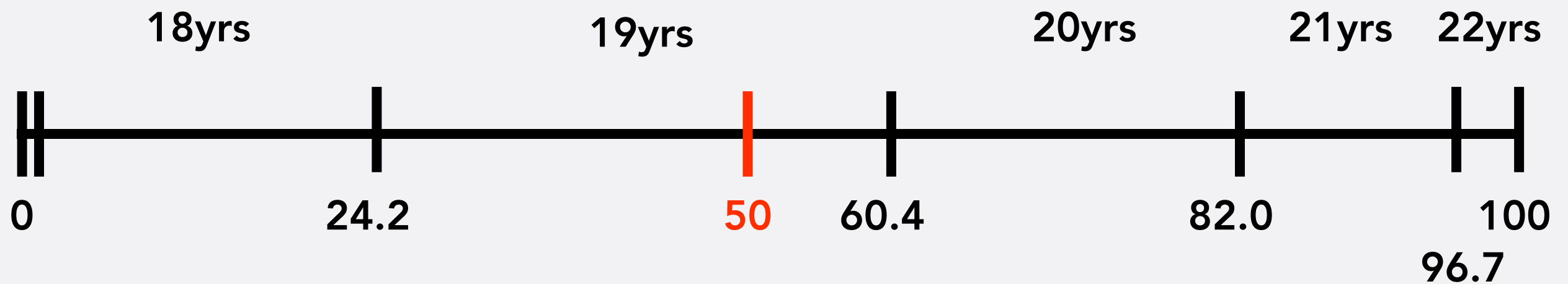
- Median: ?

FREQUENCY TABLE

Age	Number	Percentage	Cumulative Percentage
17	1	0.9	0.9
18	27	23.3	24.2
19	42	36.2	60.4
20	25	21.6	82.0
21	17	14.7	96.7
22	4	3.4	100.1

Median

- **Mode: 19**

FREQUENCY TABLE

Age	Number	Percentage	Cumulative Percentage
17	1	0.9	0.9
18	27	23.3	24.2
19	42	36.2	60.4
20	25	21.6	82.0
21	17	14.7	96.7
22	4	3.4	100.1

- **Mean: ?**

FREQUENCY TABLE

Age	Number	Percentage	Cumulative Percentage
17	1	0.9	0.9
18	27	23.3	24.2
19	42	36.2	60.4
20	25	21.6	82.0
21	17	14.7	96.7
22	4	3.4	100.1

- **Mean: 1*17**

FREQUENCY TABLE

Age	Number	Percentage	Cumulative Percentage
17	1	0.9	0.9
18	27	23.3	24.2
19	42	36.2	60.4
20	25	21.6	82.0
21	17	14.7	96.7
22	4	3.4	100.1

- **Mean: 1*17+27*18**

FREQUENCY TABLE

Age	Number	Percentage	Cumulative Percentage
17	1	0.9	0.9
18	27	23.3	24.2
19	42	36.2	60.4
20	25	21.6	82.0
21	17	14.7	96.7
22	4	3.4	100.1

- Mean: 1*17+27*18+42*19

FREQUENCY TABLE

Age	Number	Percentage	Cumulative Percentage
17	1	0.9	0.9
18	27	23.3	24.2
19	42	36.2	60.4
20	25	21.6	82.0
21	17	14.7	96.7
22	4	3.4	100.1

- Mean: 1*17+27*18+42*19+25*20

FREQUENCY TABLE

Age	Number	Percentage	Cumulative Percentage
17	1	0.9	0.9
18	27	23.3	24.2
19	42	36.2	60.4
20	25	21.6	82.0
21	17	14.7	96.7
22	4	3.4	100.1

- **Mean: 1*17+27*18+42*19+25*20+17*21**

FREQUENCY TABLE

Age	Number	Percentage	Cumulative Percentage
17	1	0.9	0.9
18	27	23.3	24.2
19	42	36.2	60.4
20	25	21.6	82.0
21	17	14.7	96.7
22	4	3.4	100.1

- **Mean: 1*17+27*18+42*19+25*20+17*21+4*22**

FREQUENCY TABLE

Age	Number	Percentage	Cumulative Percentage
17	1	0.9	0.9
18	27	23.3	24.2
19	42	36.2	60.4
20	25	21.6	82.0
21	17	14.7	96.7
22	4	3.4	100.1

- **Mean: (1*17+27*18+42*19+25*20+17*21+4*22)/116=19.4**

SLEEP

- **How many hours to you sleep at night?**
  - **Regular week:**
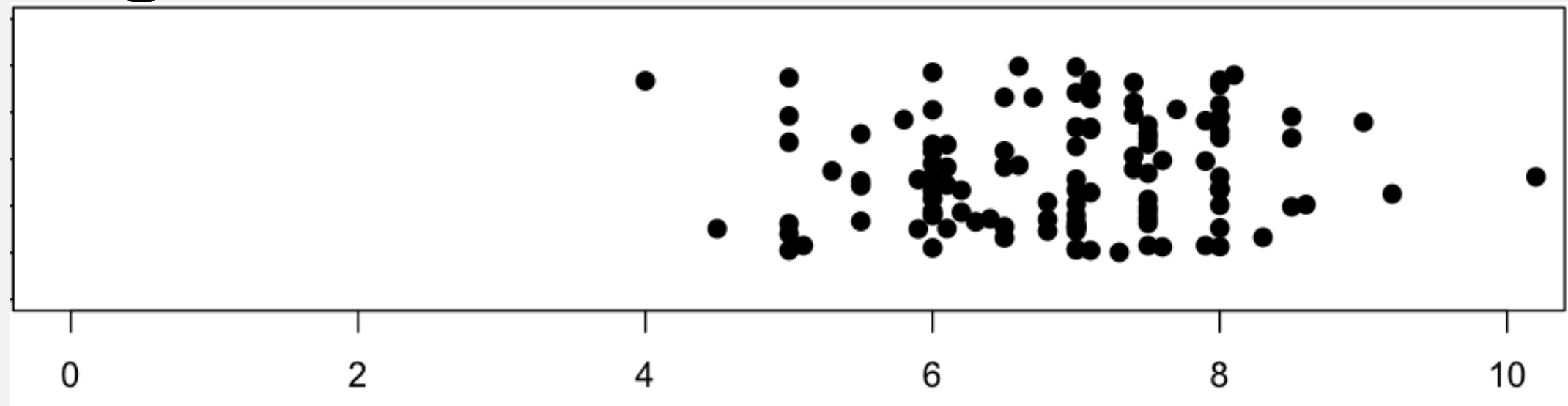
  - **Finals week:**

SLEEP

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

Mean: 6.9

  - **Finals week:**

Mean: 5.8
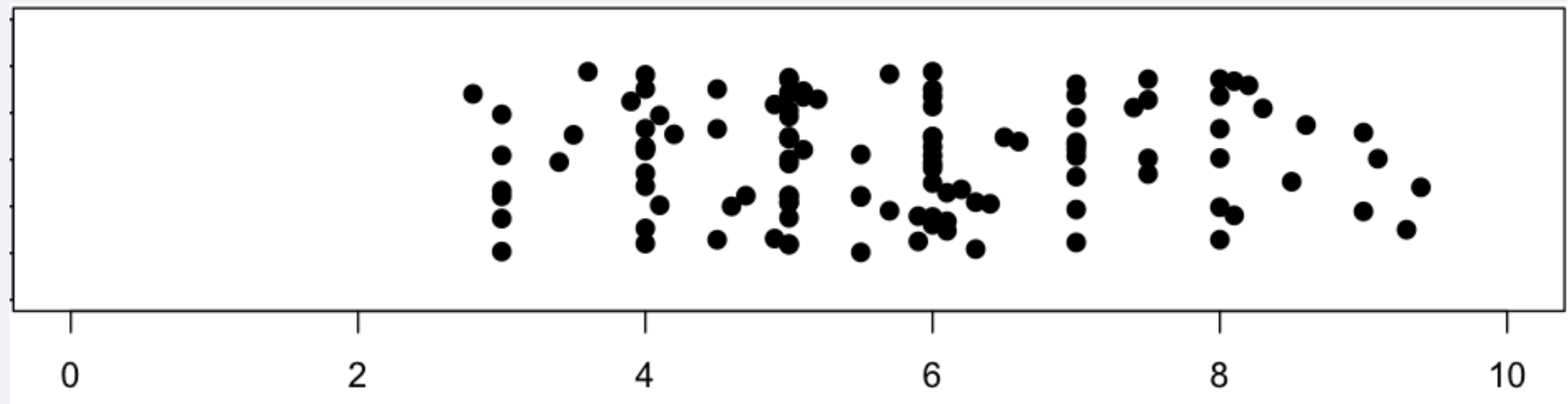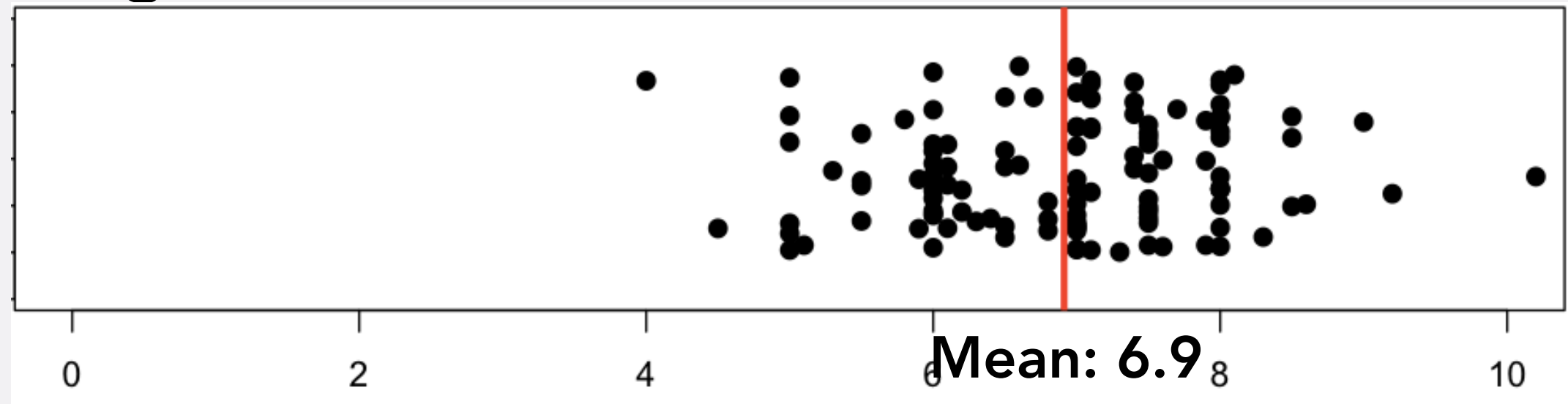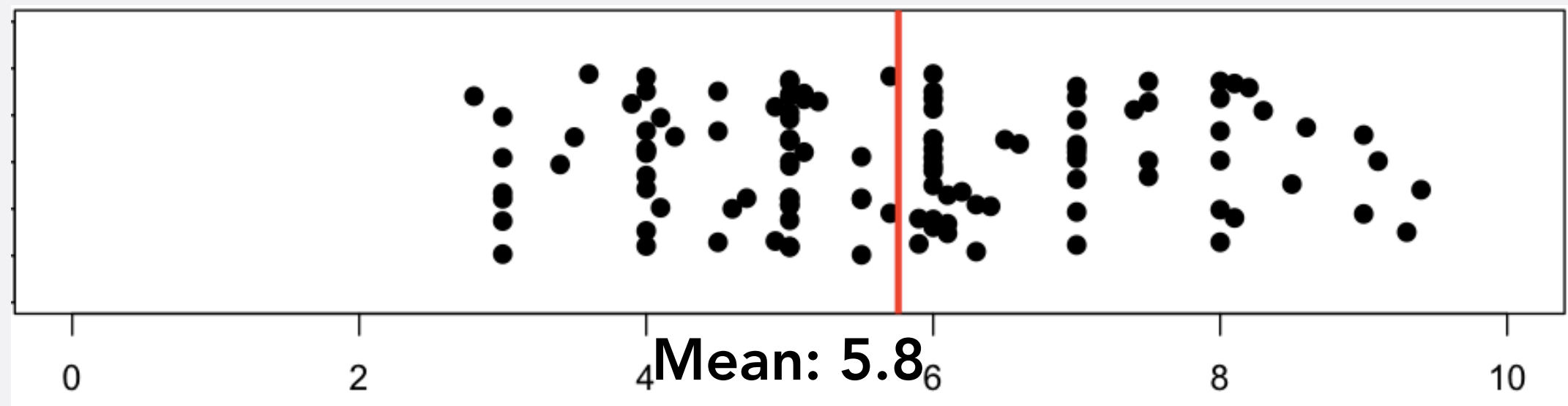
DISPERSION

- **How many hours to you sleep at night?**
  - **Regular week:**
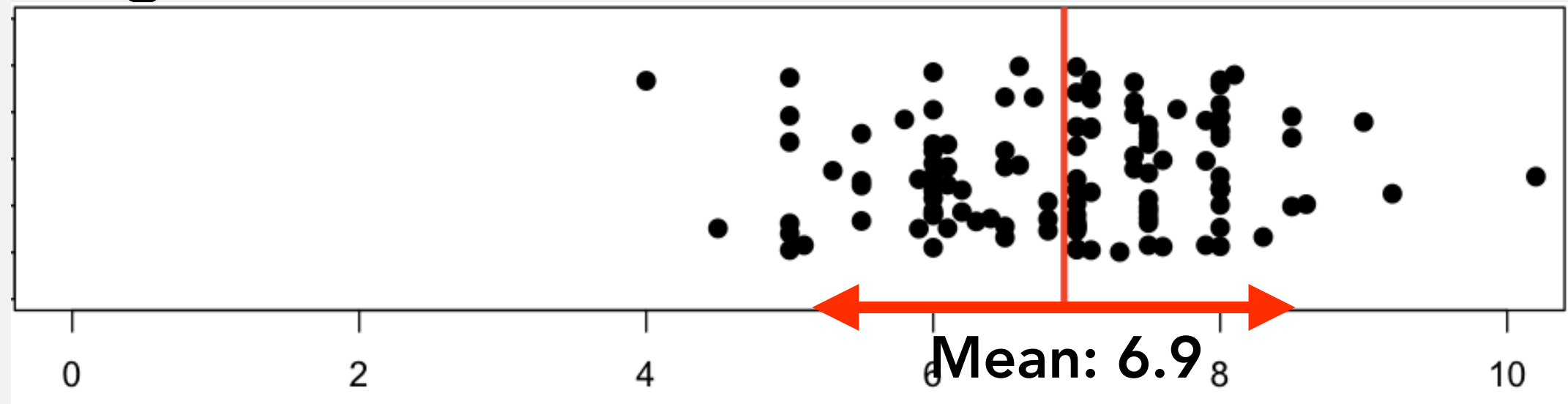
  - **Finals week:**
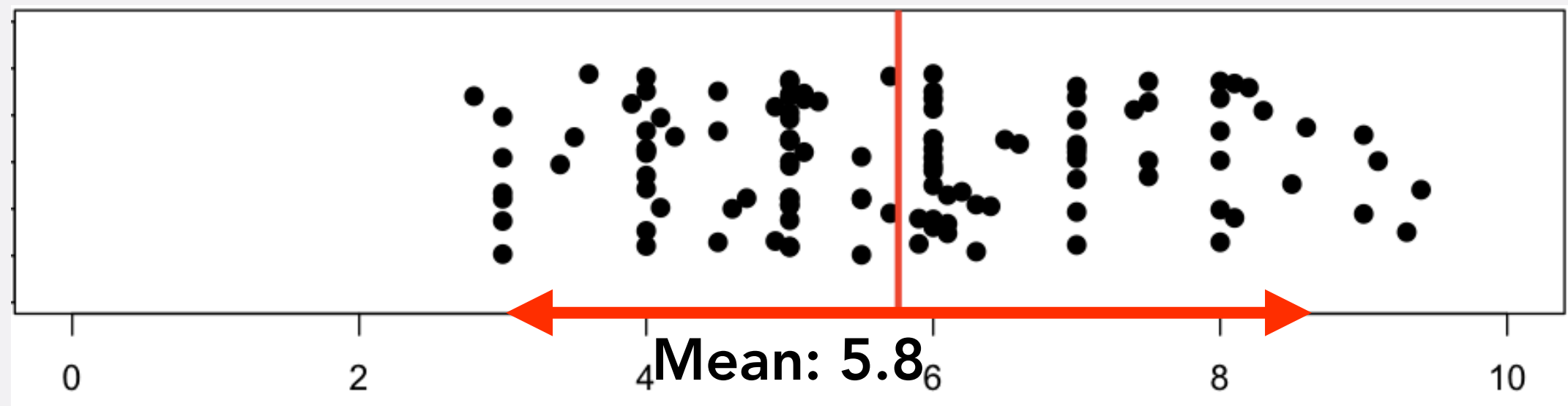
DISPERSION

- ## All three: Median 7, Mean 7
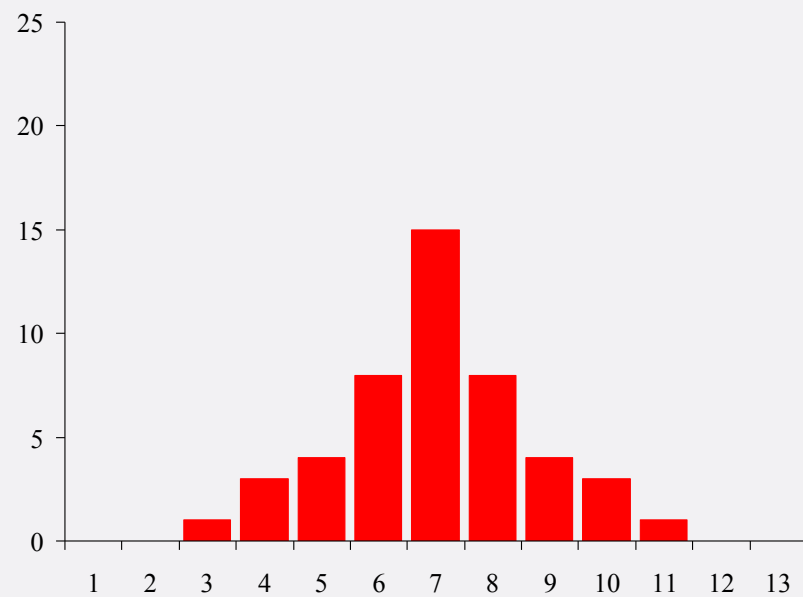
Many students perform mediocre, some do well, others not
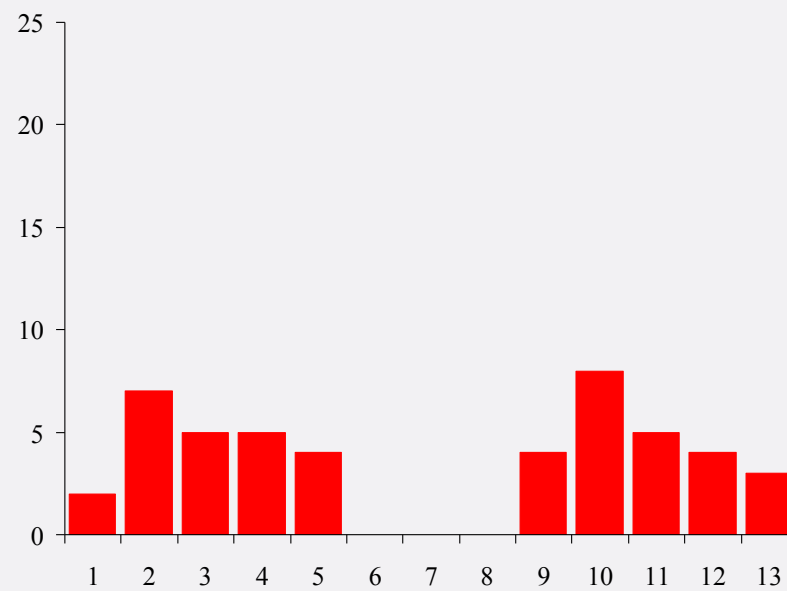
One group does very well, one group does not
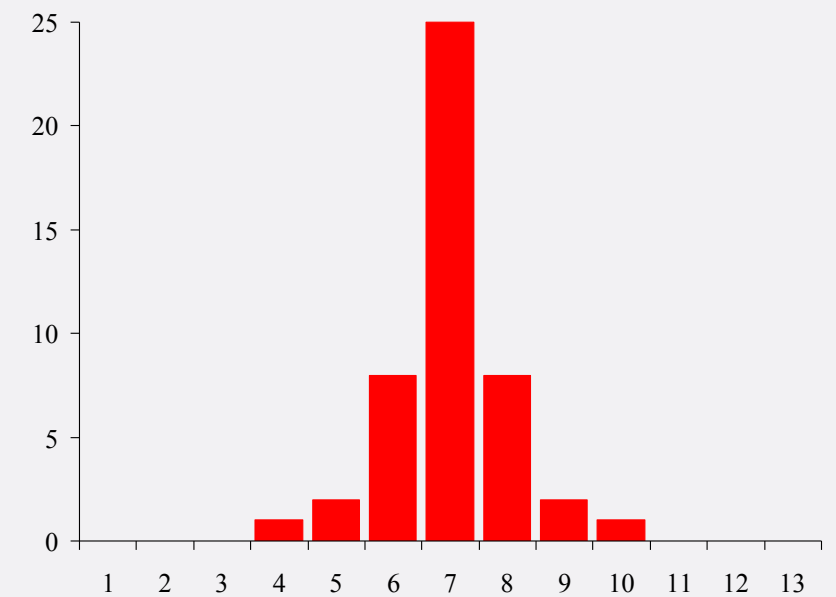
All students perform relatively similarly (mediocre)

DESCRIBING VARIABLES

- **Central tendency**
- **Frequency tables**
- Dispersion

DISPERSION

- ## Want to look at degree of variation around central tendency
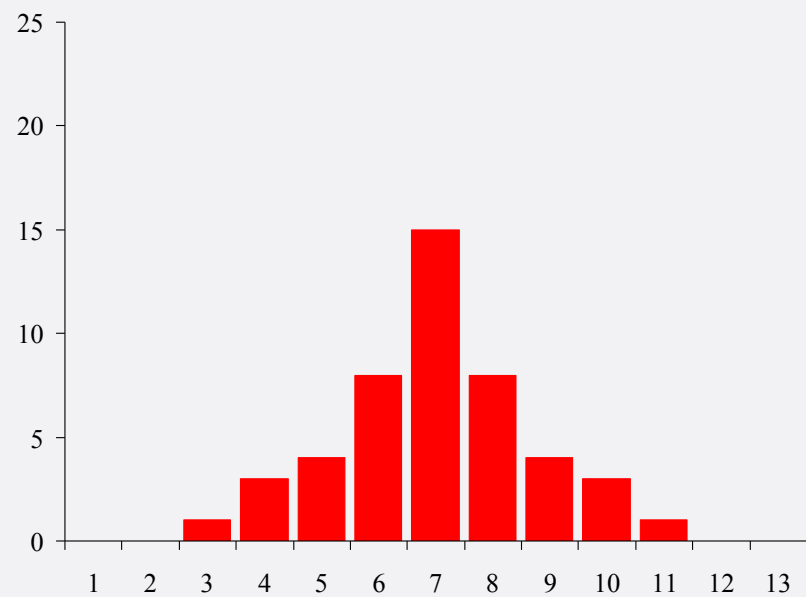
Moderate variation
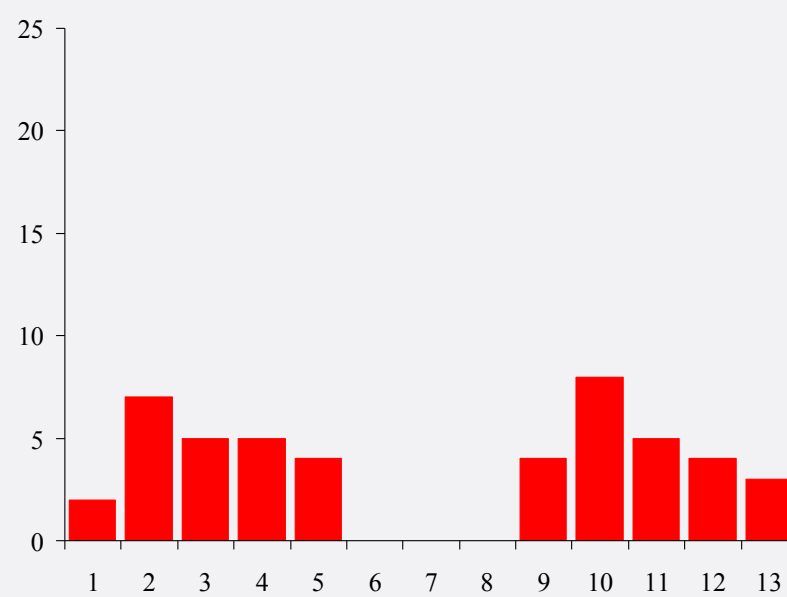
High variation
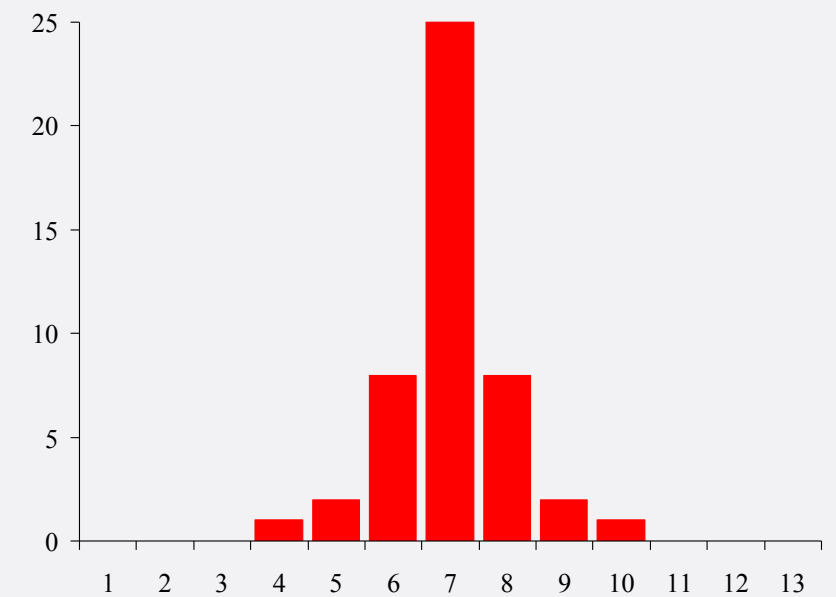
Low variation

DISPERSION

- **Want to look at degree of variation around central tendency**
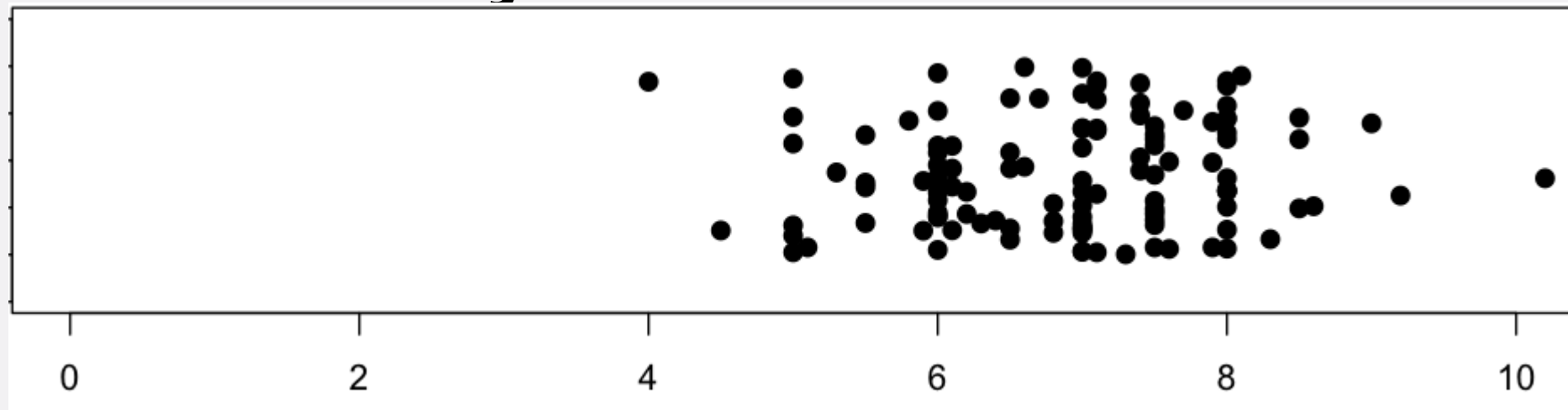
Low variation
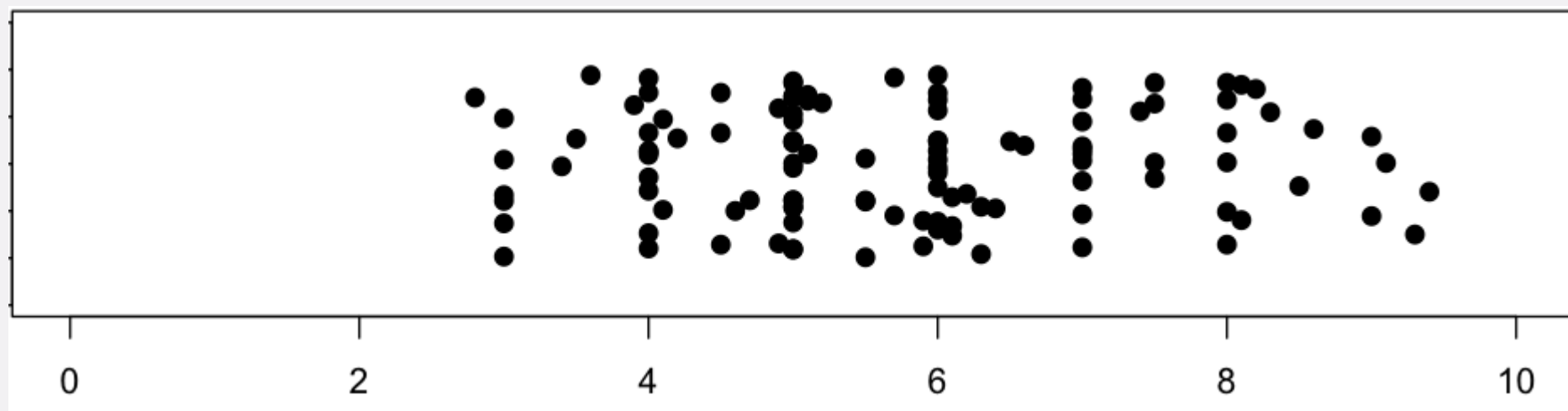
High variation

- **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**
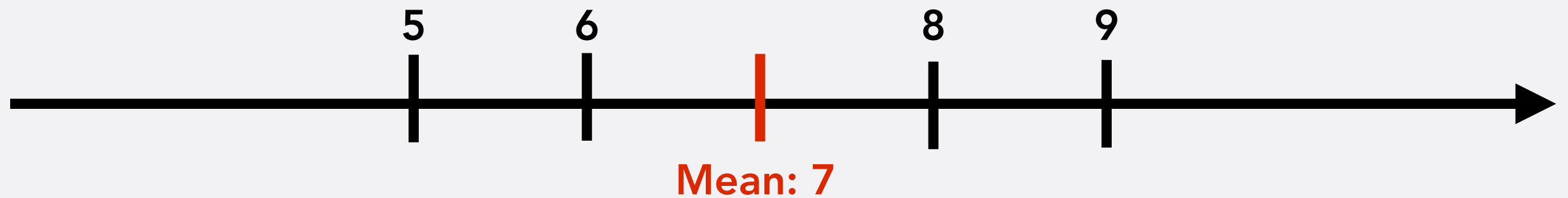
DISPERSION

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

5   6 | 8   9

Mean: 7

←———

1
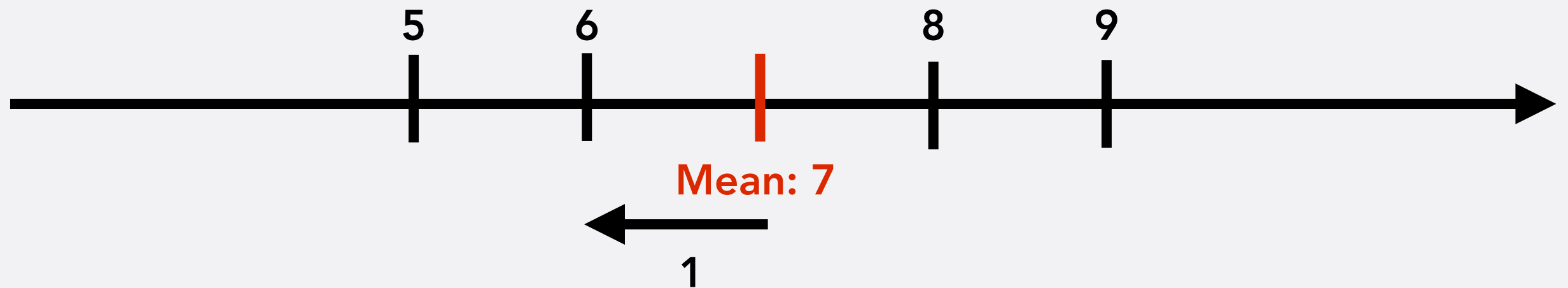
DISPERSION

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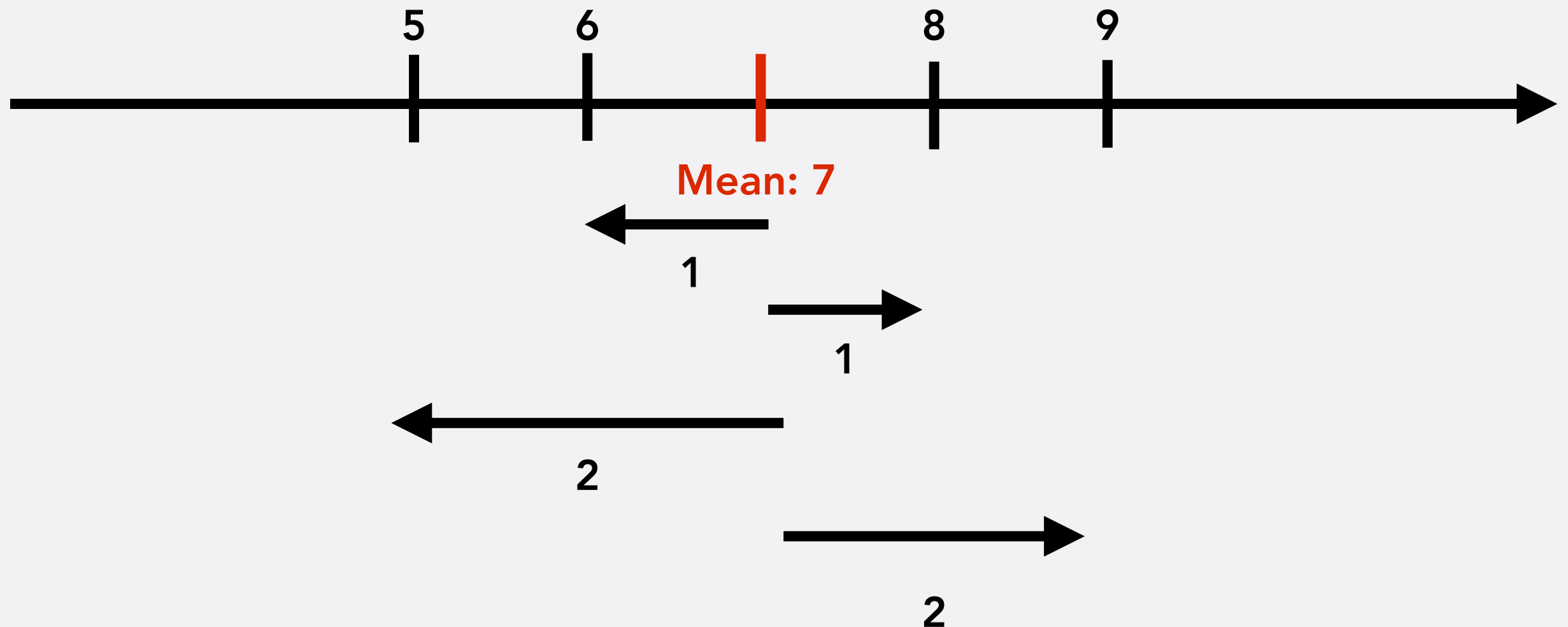
DISPERSION

- **Simpler sleep example**
  - **4 people: 5, 6, 8, 9 hours/night**
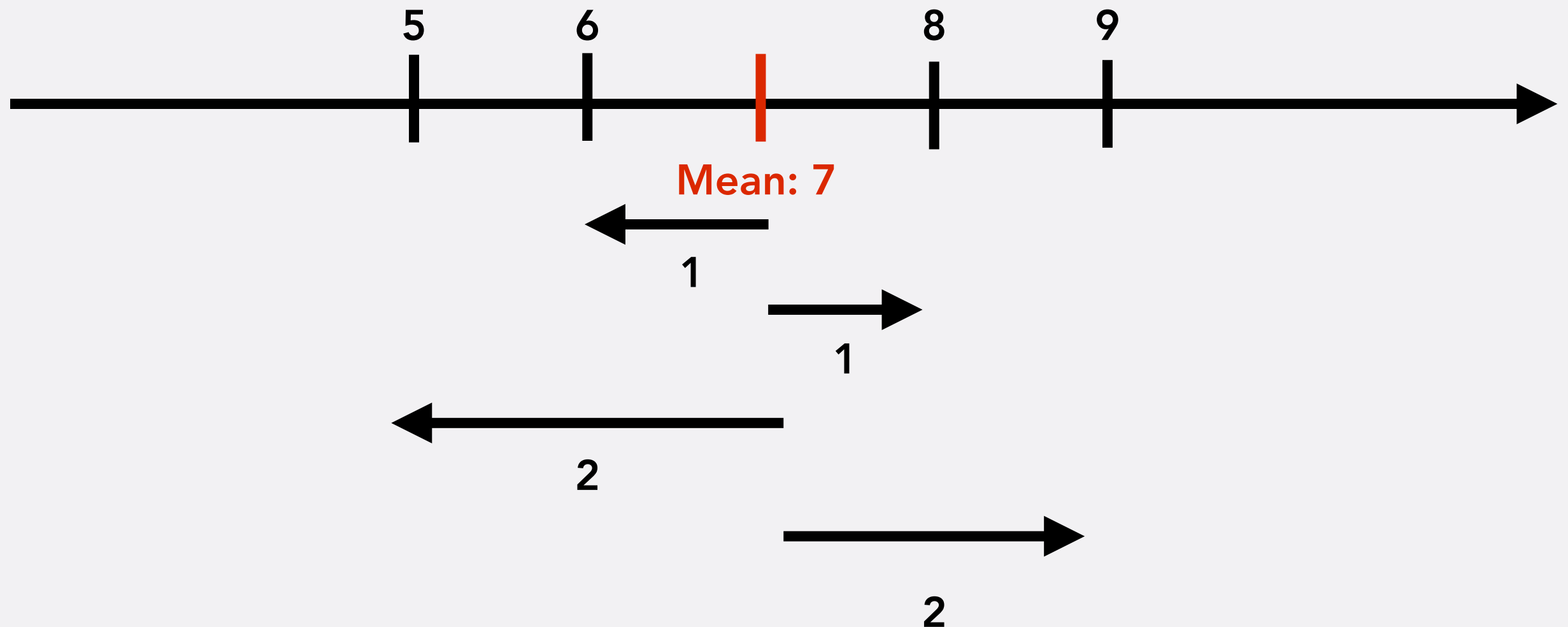
- **Average distance from mean: (1+1+2+2)/4=1.5**

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
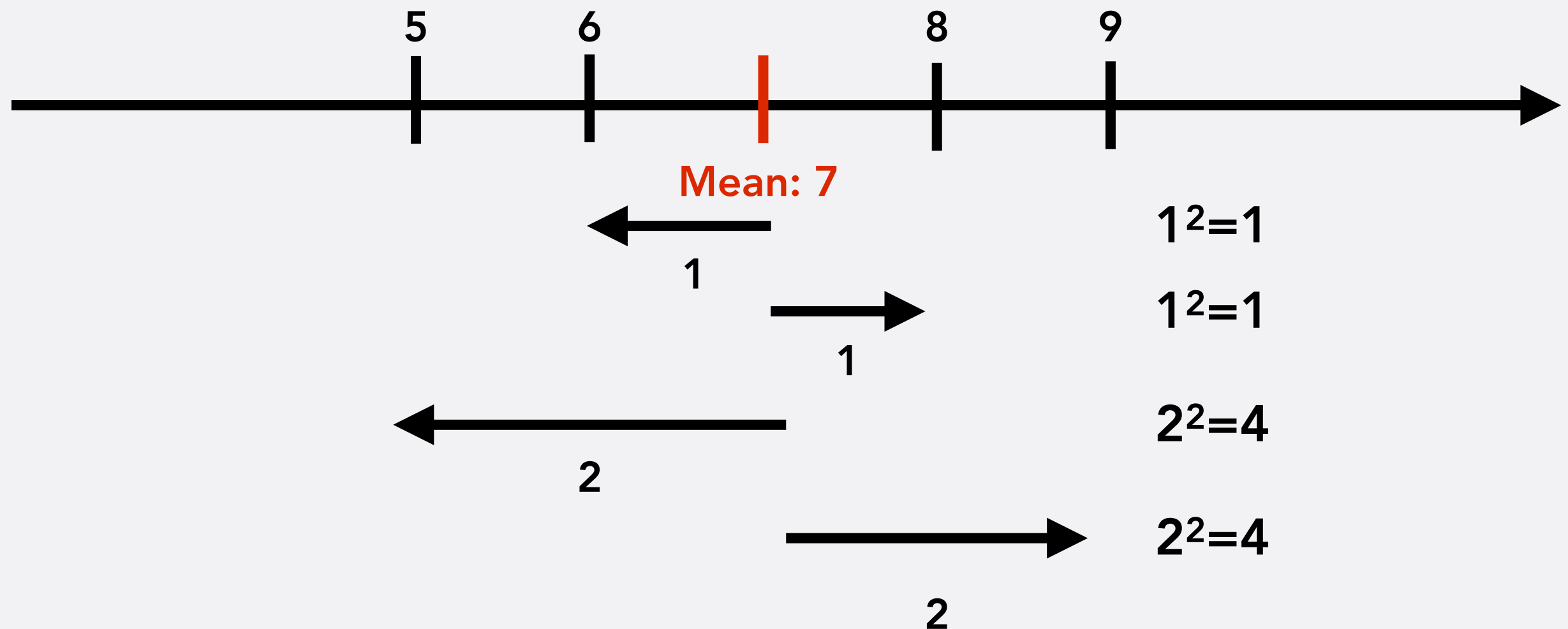 - Gives extreme cases (far from mean) more weight

STANDARD DEVIATION

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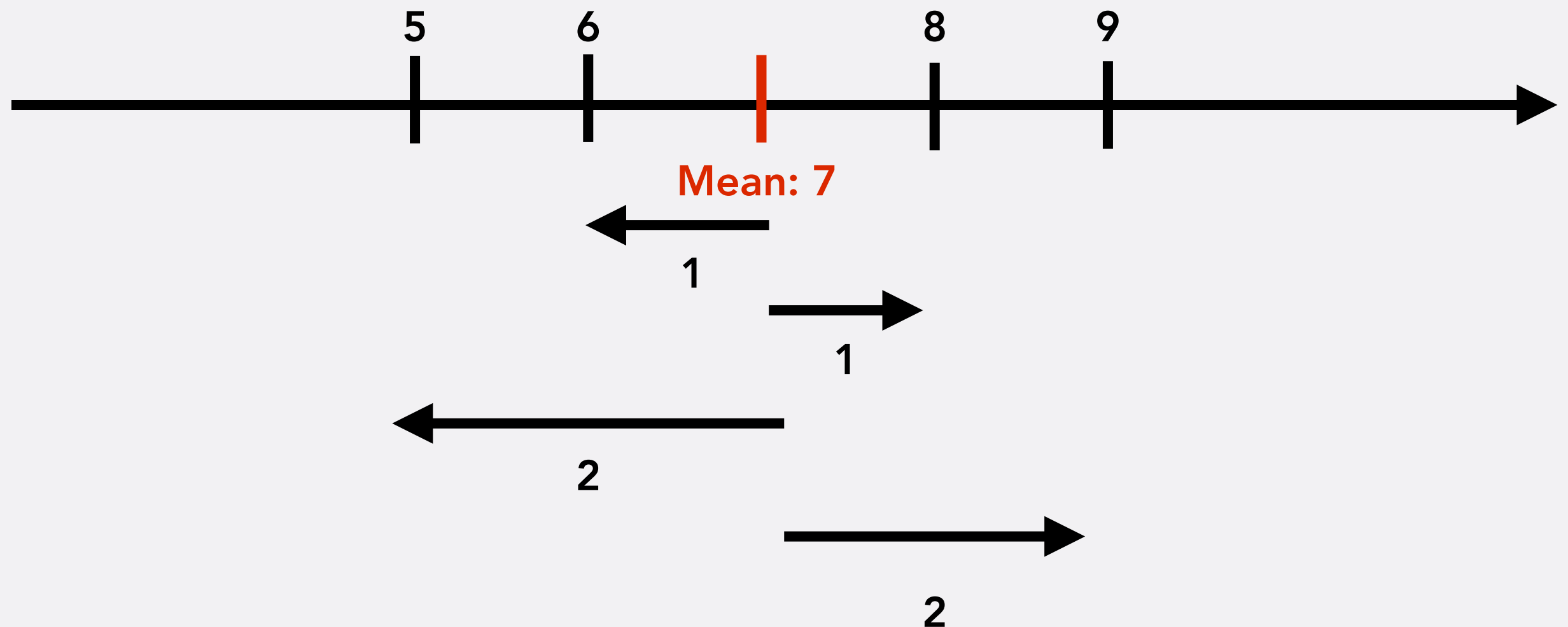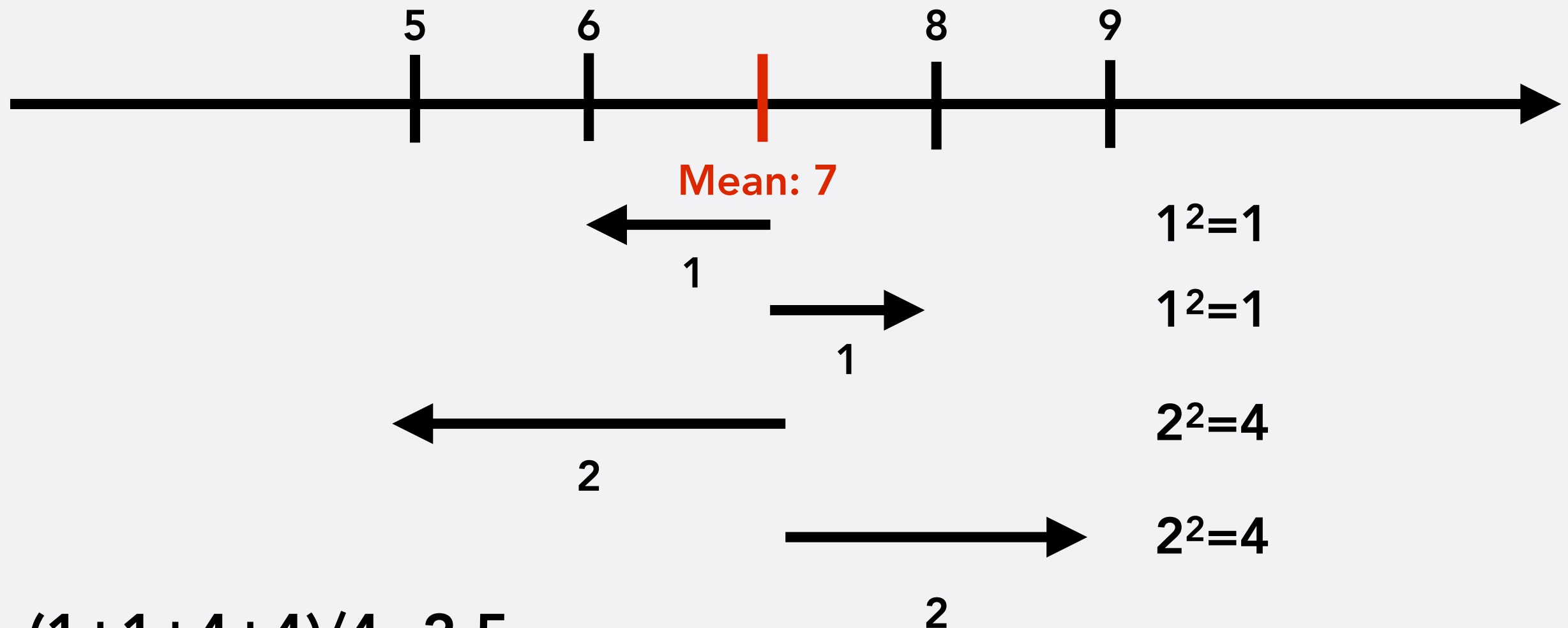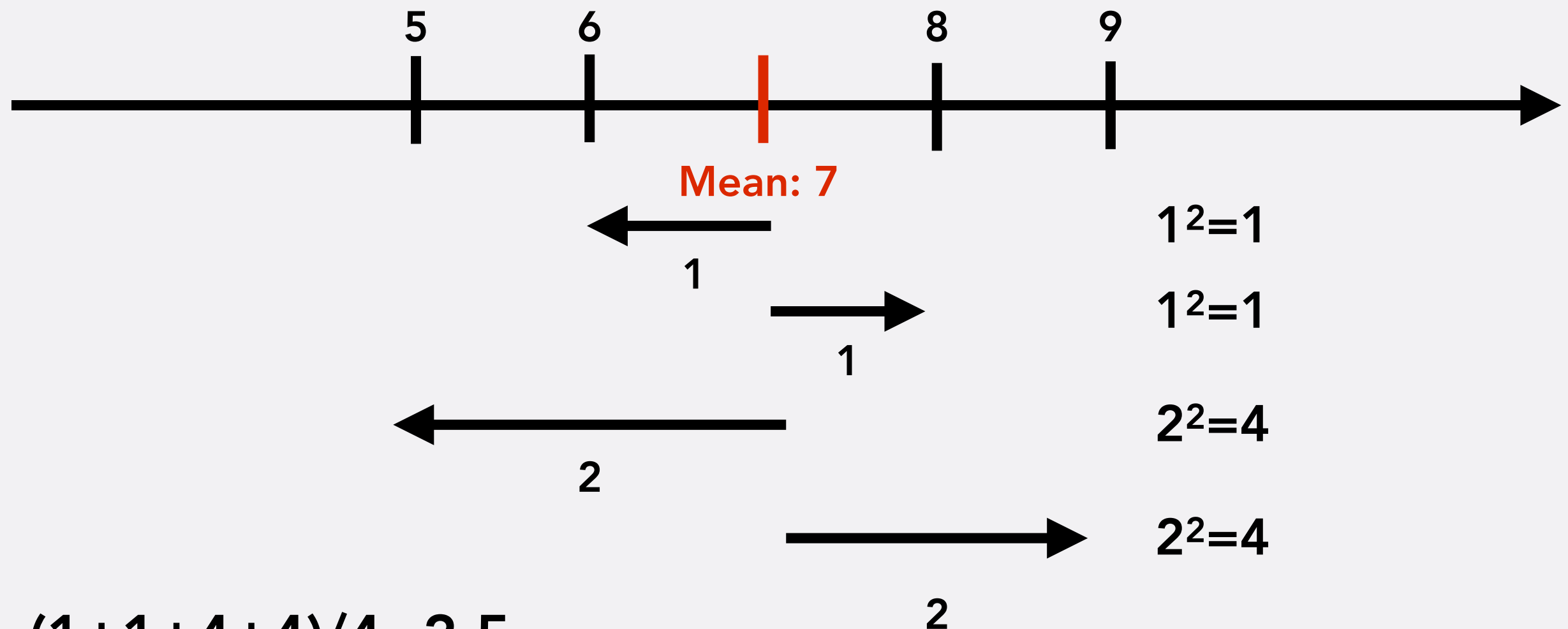
STANDARD DEVIATION

- **But now we square each of them**

| 5 | 6 | | 8 | 9 |

Mean: 7

$1^2=1$

$1^2=1$

$2^2=4$

$2^2=4$

1

1

2

2

STANDARD DEVIATION

- **Then take the average of those squared deviations**

5   6   8   9

Mean: 7

$1^2=1$

1

$1^2=1$

1

$2^2=4$

2

$2^2=4$

2

- **(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**

5   6   8   9

Mean: 7

$1^2=1$

$1^2=1$

1

$2^2=4$

2

$2^2=4$

2

- **(1+1+4+4)/4=2.5**
- **√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")