Frequentist Hypothesis Testing: *Intuitions*

Objectives

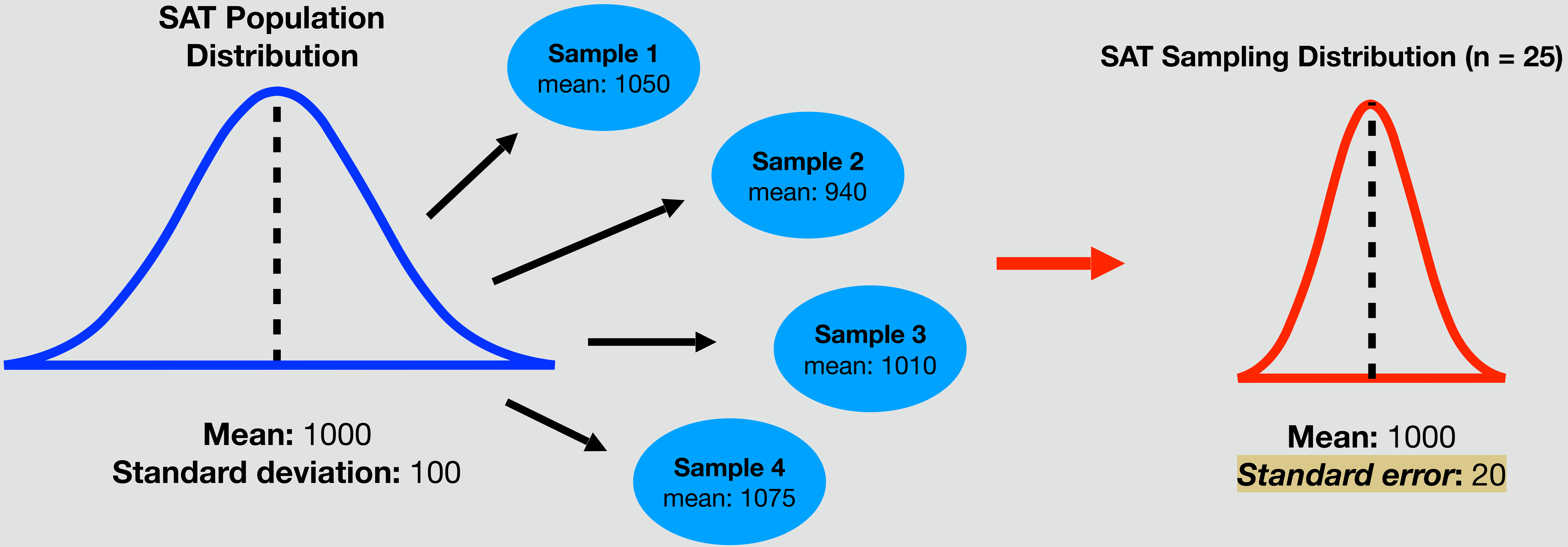
- Develop a robust intuition for the frequentist approach to hypothesis testing
- Distinguish a population distribution from a sampling distribution
- Relate the area under a sampling distribution to *P-values*
- Perform a Z-test
- Contrast the use-case of a t-test vs a Z-test

Real or random?

- The goal of hypothesis testing is to determine if an effect is ***statistically significant*** based on data
- **Example:**
  - Do SmartCards™ product actually improve SAT scores?
  - Take a random sample of 25 students, have them study with the cards:
    - Population mean (without SmartCards) : 1000
    - Experimental mean: 1050
- **Can we quantify our certainty that this effect isn't due to randomness?**

Sampling Distributions

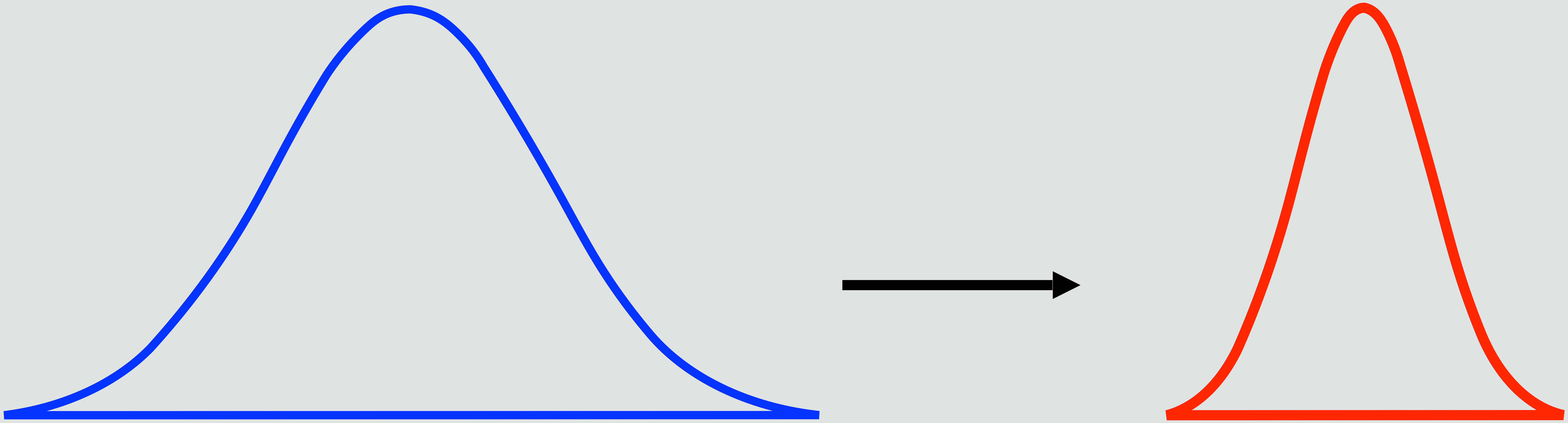
~~ Draw 25 person samples ~~



Peculiar Population Distributions

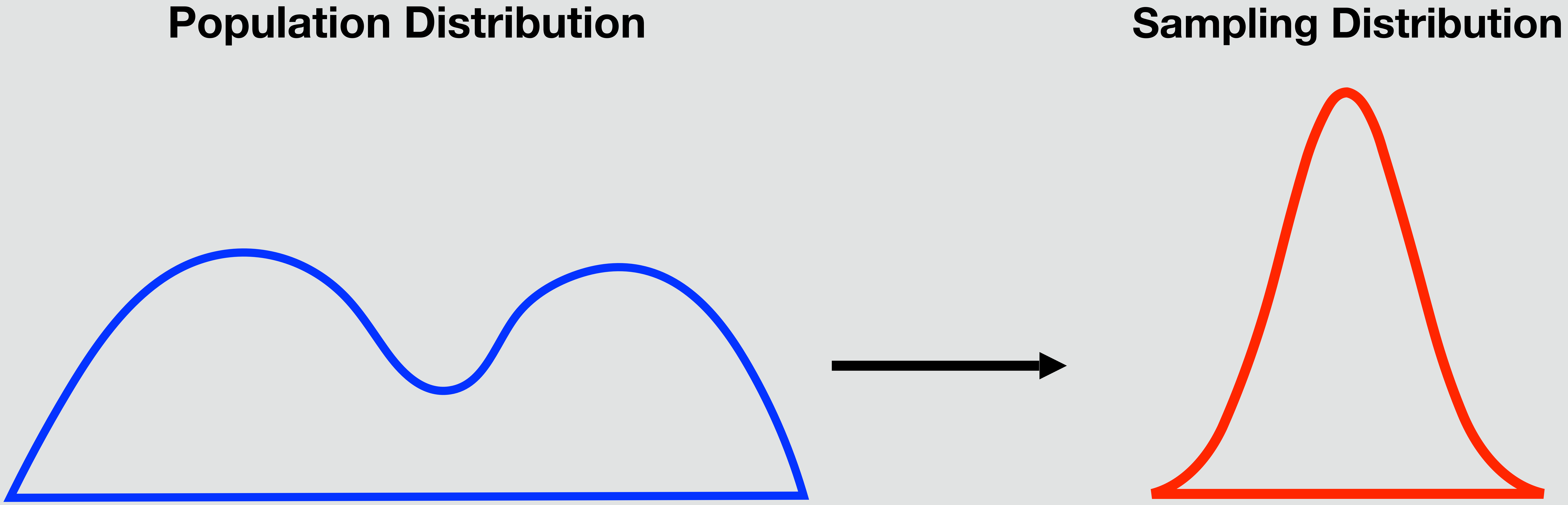
Nasty population distributions still have nice and normal sampling distributions

Population Distribution **Sampling Distribution**



Peculiar Population Distributions

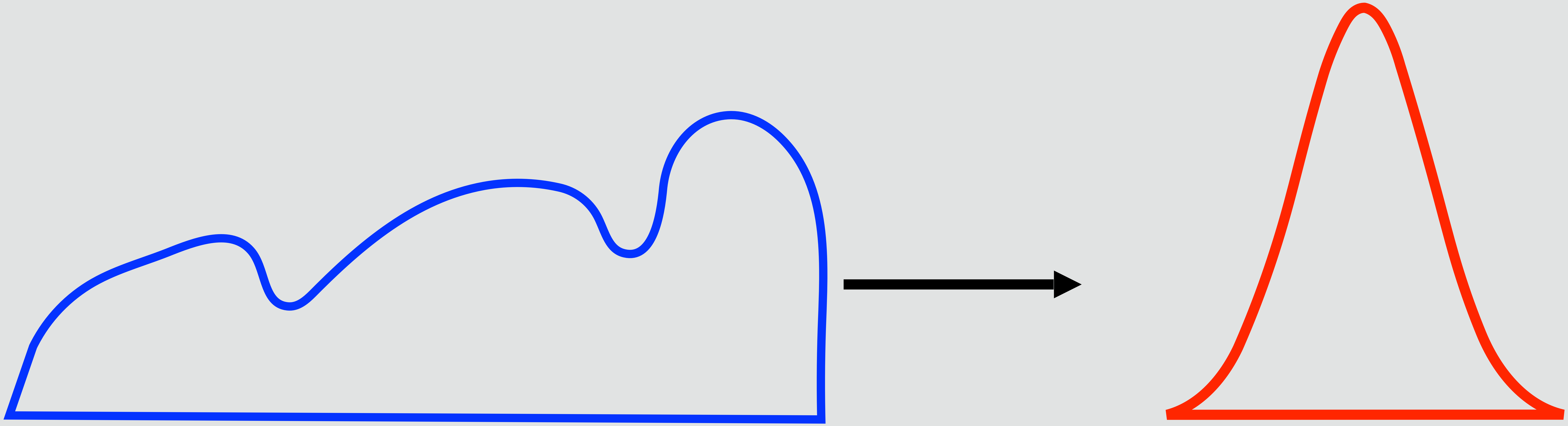
Nasty population distributions still have nice and normal sampling distributions



Peculiar Population Distributions

Nasty population distributions still have nice and normal sampling distributions

Population Distribution **Sampling Distribution**



Peculiar Population Distributions

Nasty population distributions still have nice and normal sampling distributions

Who do we have to thank for this?

THE CENTRAL LIMIT THEOREM

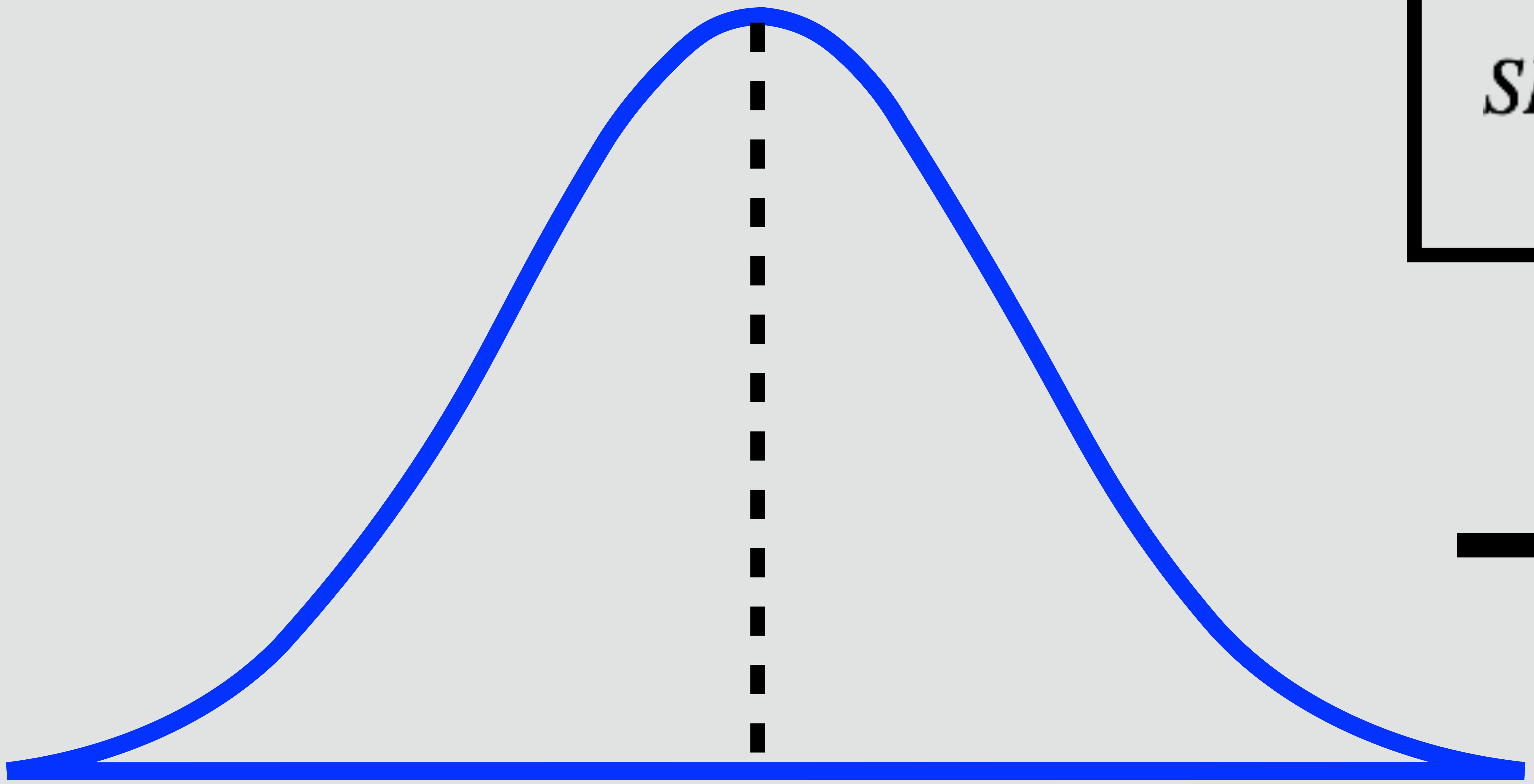
Our friend, the standard error

- The standard deviation of sample means
- Measures the stability of sample means
- **Depends on:**
    - The population standard deviation
    - Sample size

$$SE = \frac{\sigma}{\sqrt{n}}$$

Sample Size affects the sampling distribution

SAT Population Distribution

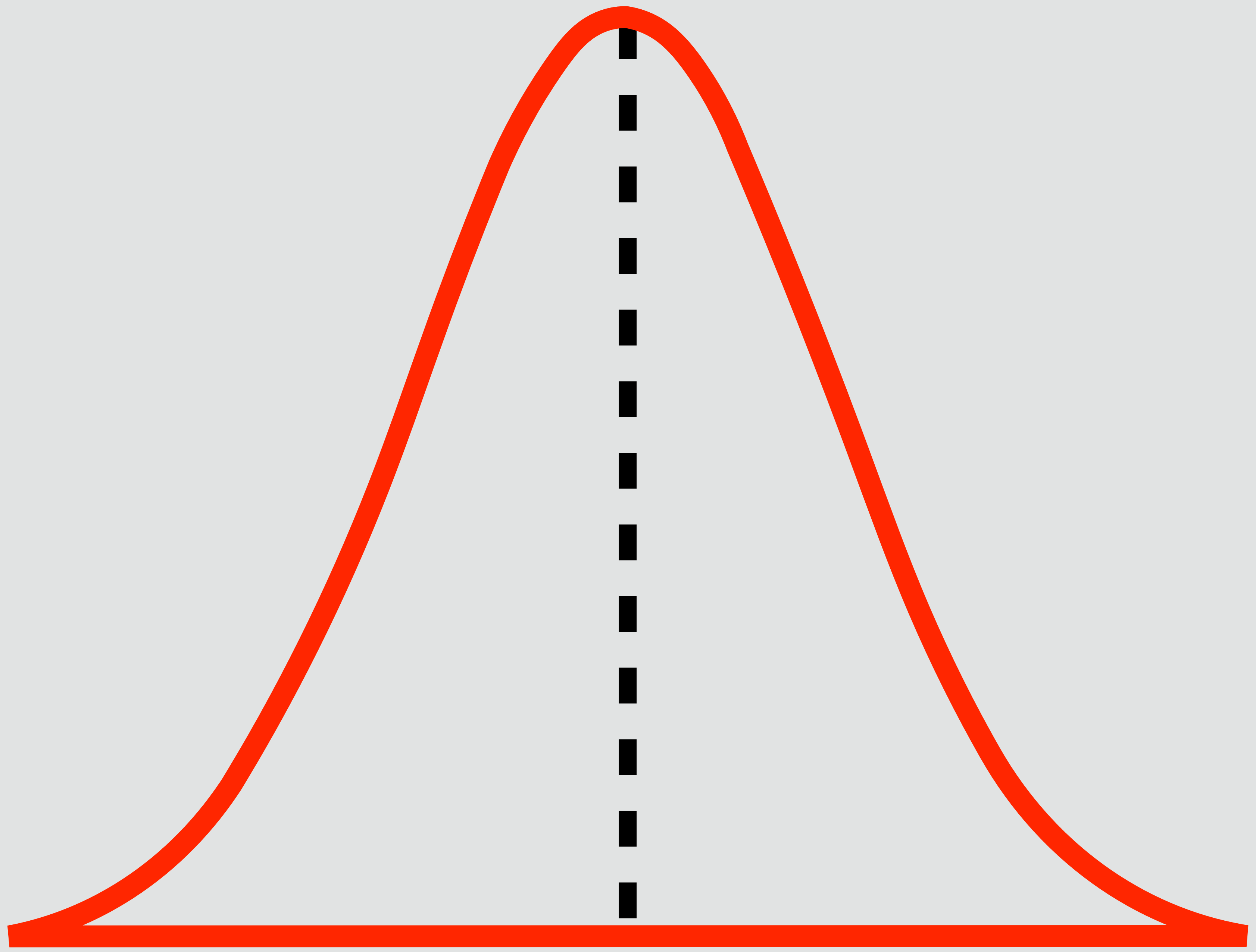


Mean: 1000
Standard deviation: 100

$$SE = \frac{100}{\sqrt{4}}$$

n = 4

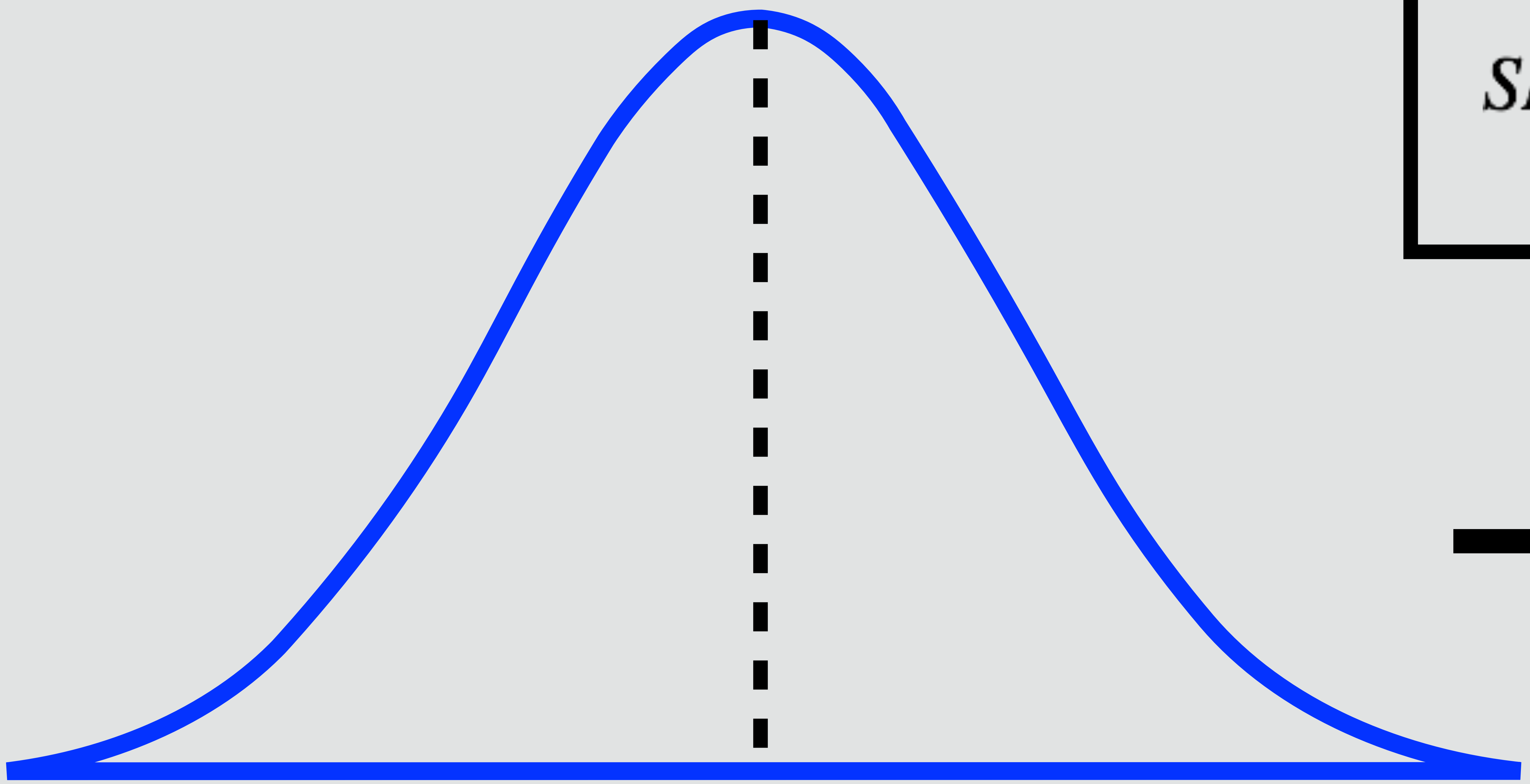
SAT Sampling Distribution (n = 4)



Mean: 1000
***Standard error*: 50**

Sample Size affects the sampling distribution

SAT Population Distribution

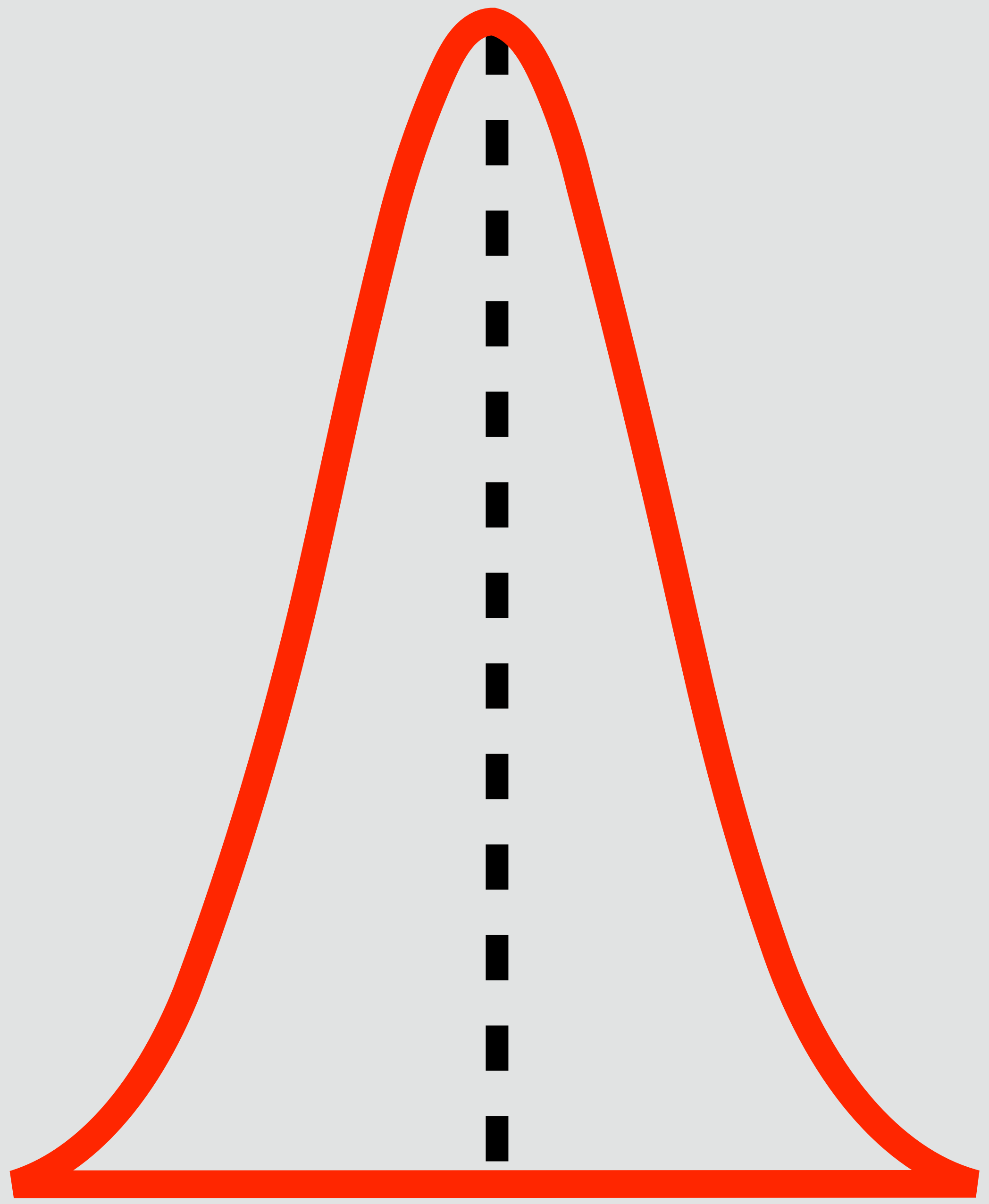


Mean: 1000
Standard deviation: 100

$$\boxed{SE = \frac{100}{\sqrt{25}}}$$

n = 25

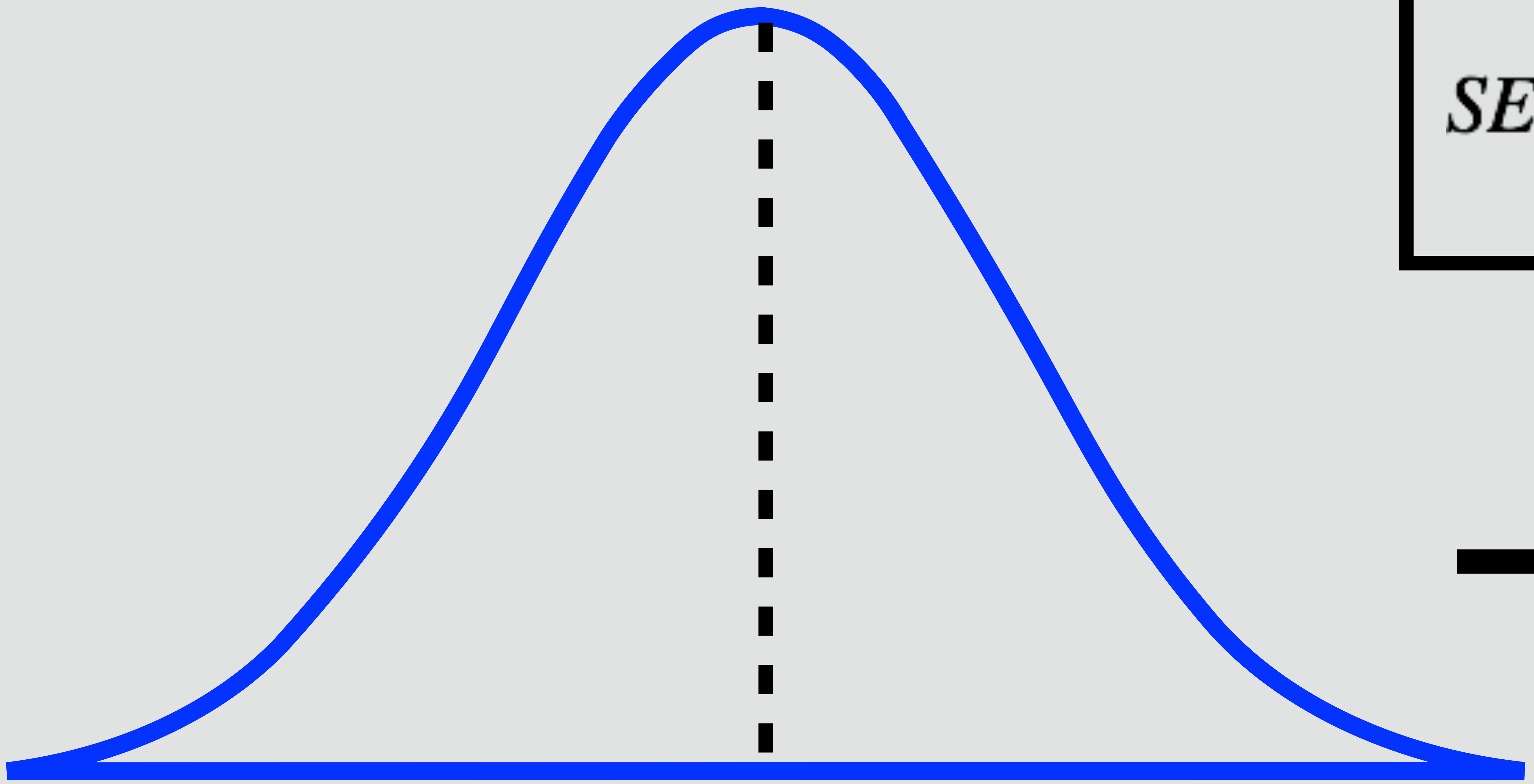
SAT Sampling Distribution (n = 25)



Mean: 1000
***Standard error*: <u>*20*</u>**

Sample Size affects the sampling distribution

SAT Population Distribution

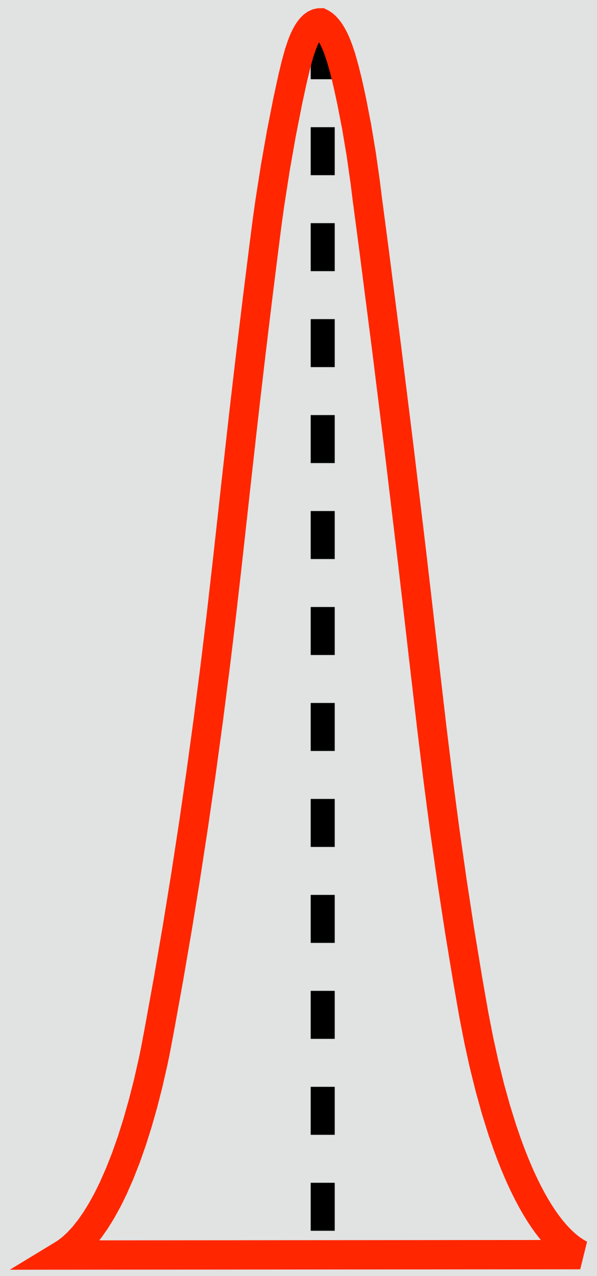


Mean: 1000
Standard deviation: 100

$$SE = \frac{100}{\sqrt{100}}$$

n = 100

SAT Sampling Distribution (n = 100)

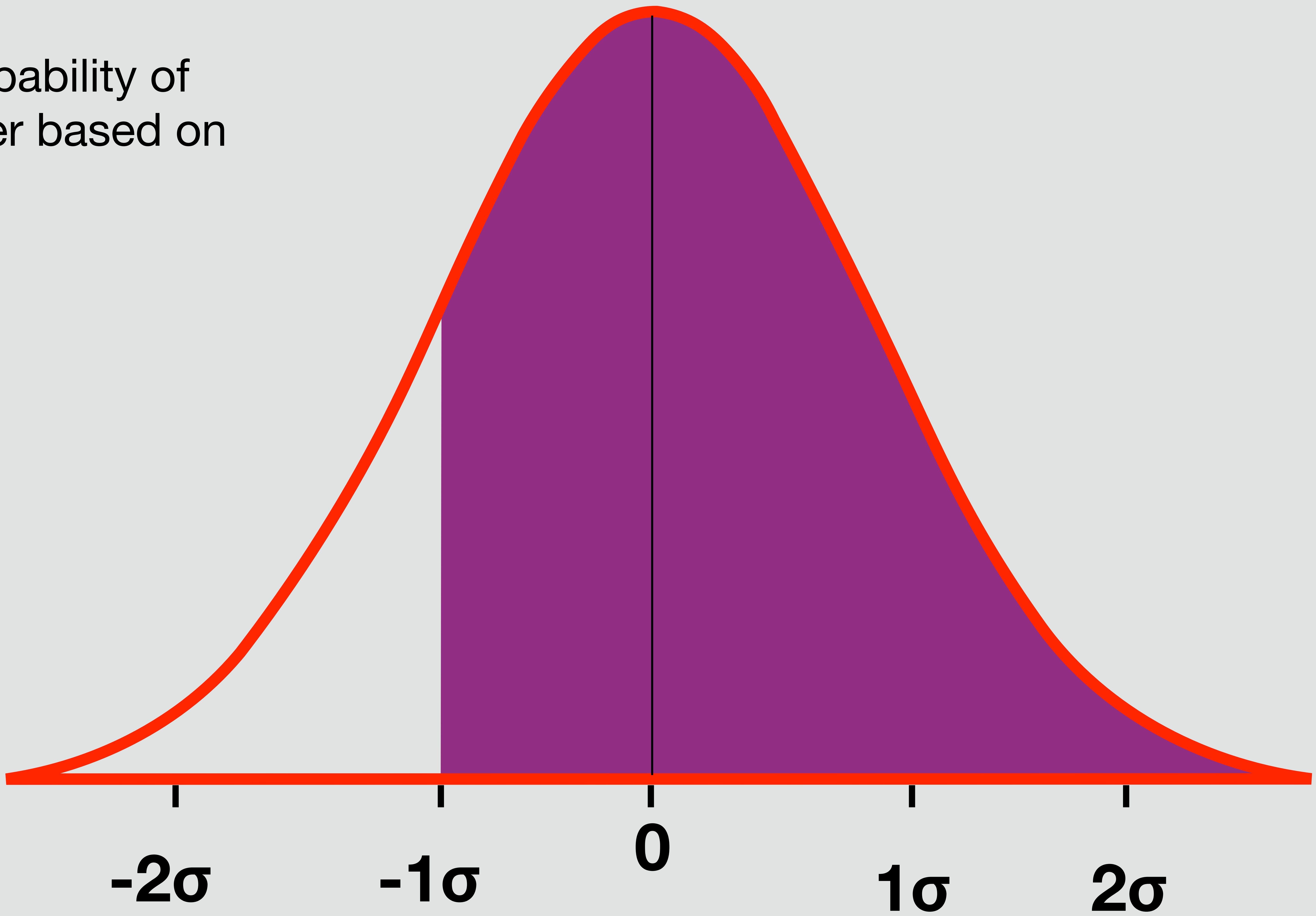


Mean: 1000
Standard error*: *10

Probability <> area under the curve

- We can calculate the probability of drawing a mean or greater based on area under the curve

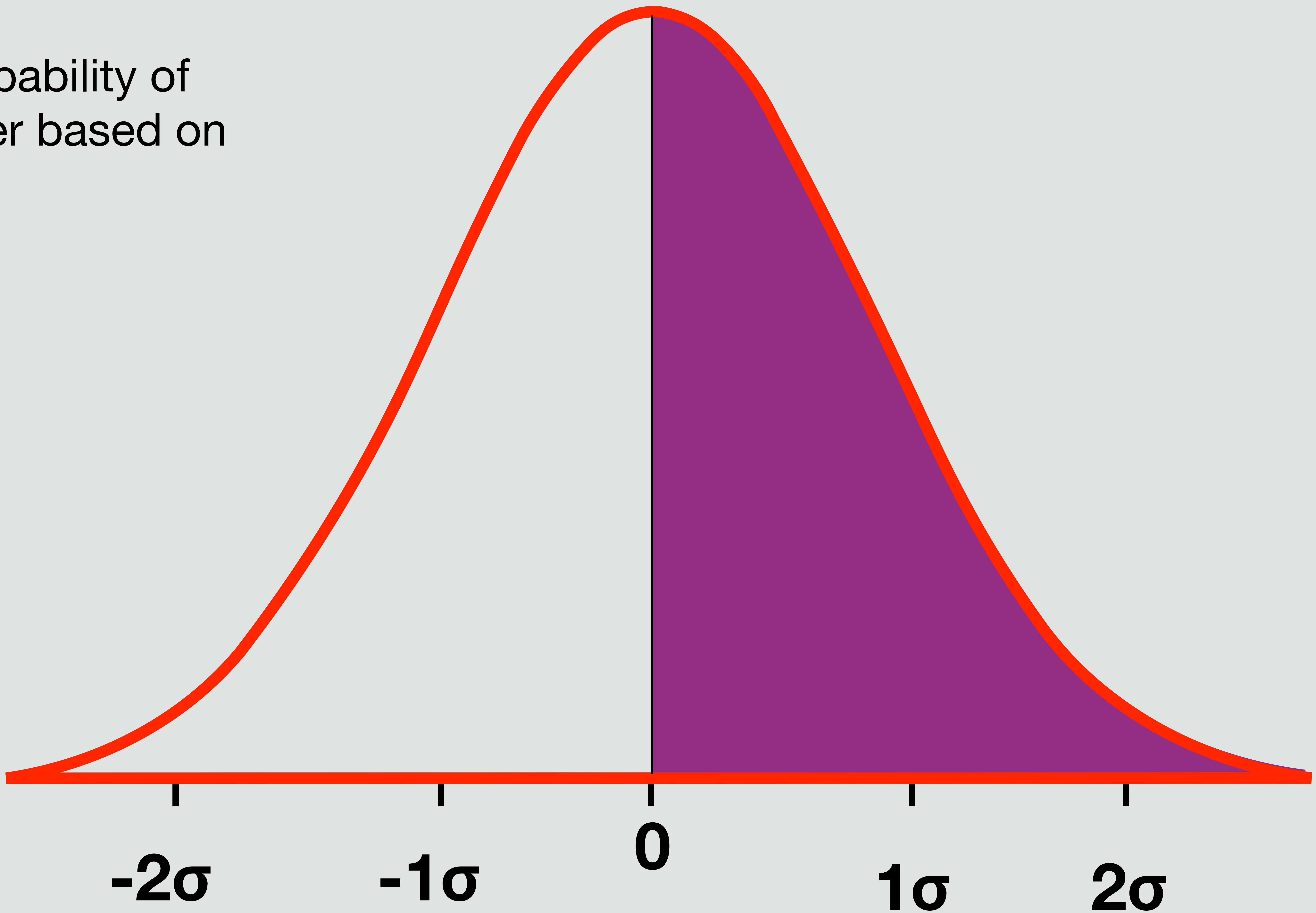
P = 0.85



Probability <> area under the curve

- We can calculate the probability of drawing a mean or greater based on area under the curve

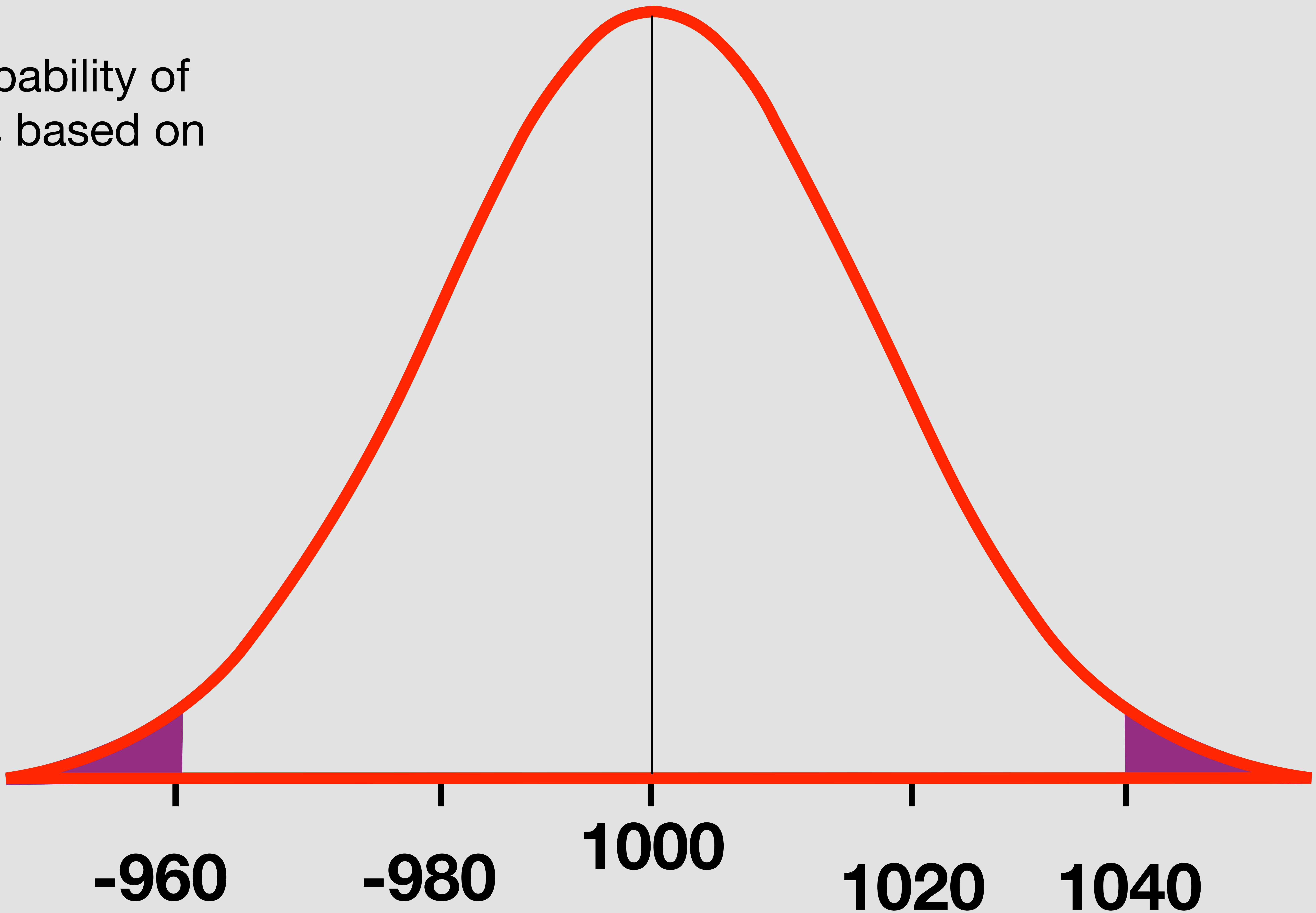
P = 0.50



Probability <> area under the curve

- We can calculate the probability of drawing ranges of means based on area under the curve

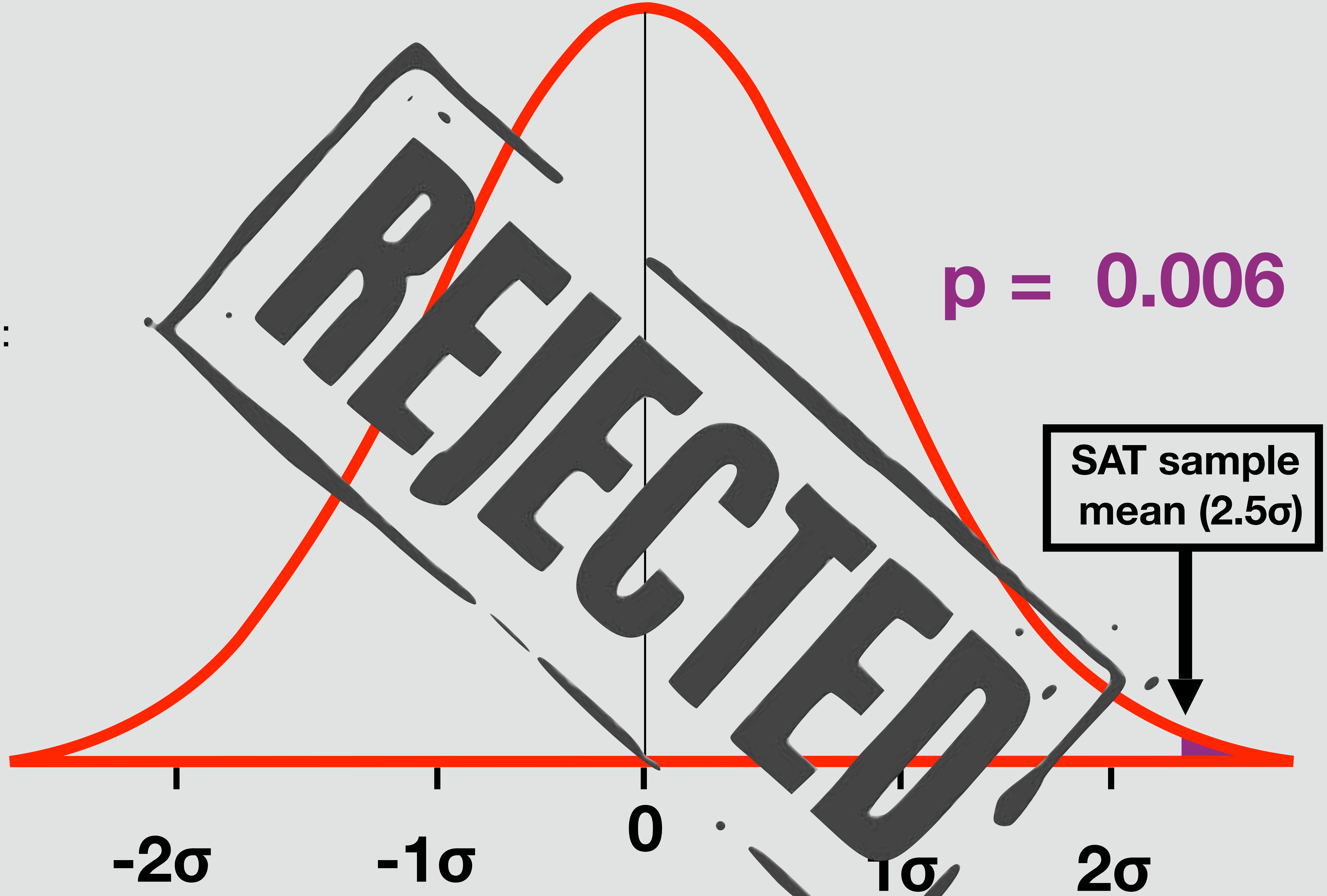
P = 0.046



Back to the SAT example

- Standard SAT population:
  - Mean: 1000
  - STD: 100
- Sampling distribution (n = 25):
  - Mean: 1000
  - Standard Error: 20
- Experimental Results
  - Mean: 1050

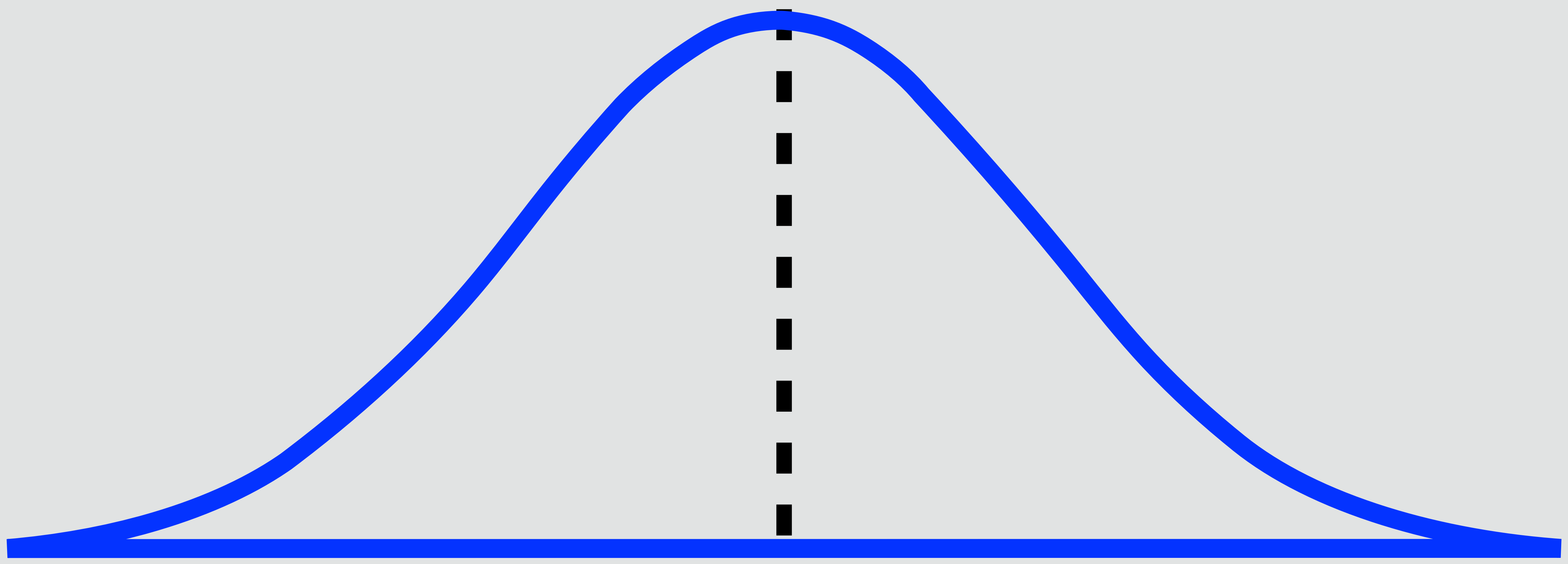
$$Z = \frac{1050 - 1000}{20}$$



Where the Z-test falters…

- We used ***known*** population parameters, mean and standard deviation to construct a sampling distribution
- It is rare to have that information

# of cartwheels in a minute



Mean: ???
Standard deviation: ???

… the t-test prevails

- We don’t know the exact population mean and standard deviation, so we conservatively estimate them based on samples themselves
- If we didn’t know SAT population parameters we might:
    - Test one random sample of 25, untreated as the control
    - Test another random sample of 25 who studied the flash cards
    - Compute test statistic and p-value based on t-distribution, as opposed to the normal (z) distribution

$$t = \frac{M_1 - M_2}{SE} \qquad SE = \sqrt{\frac{s_1^2 + s_2^2}{n}}$$

Summary

- The Central Limit Theorem makes sampling distributions predictable
- We construct a sampling distribution based on the null-hypothesis
- If it seems unlikely that the observed data came from that distribution, we reject the null hypothesis
- Use a t-test (or some other hypothesis test) when population parameters are unknown (which is most of the time). However

Further reading

- Khan on T-distribution vs Z-distribution - https://www.youtube.com/watch?v=5ABpqVSx33I
- Free Udacity course on inferential statistics - https://www.udacity.com/course/intro-to-inferential-statistics--ud201
- Wikipedia (has surprisingly good articles on these topics)
    - https://en.wikipedia.org/wiki/Sampling_distribution
    - https://en.wikipedia.org/wiki/Student%27s_t-test
    - https://en.wikipedia.org/wiki/Confidence_interval
    - https://en.wikipedia.org/wiki/Analysis_of_variance