SOC-5811 Week 5: Sampling distributions and inference I

Nick Graetz

University of Minnesota, Department of Sociology

9/29/2025

Statistical inference

- **Parameters** = the unknown numbers that determine a statistical model.
- Parameters can be used to simulate new data from the model.

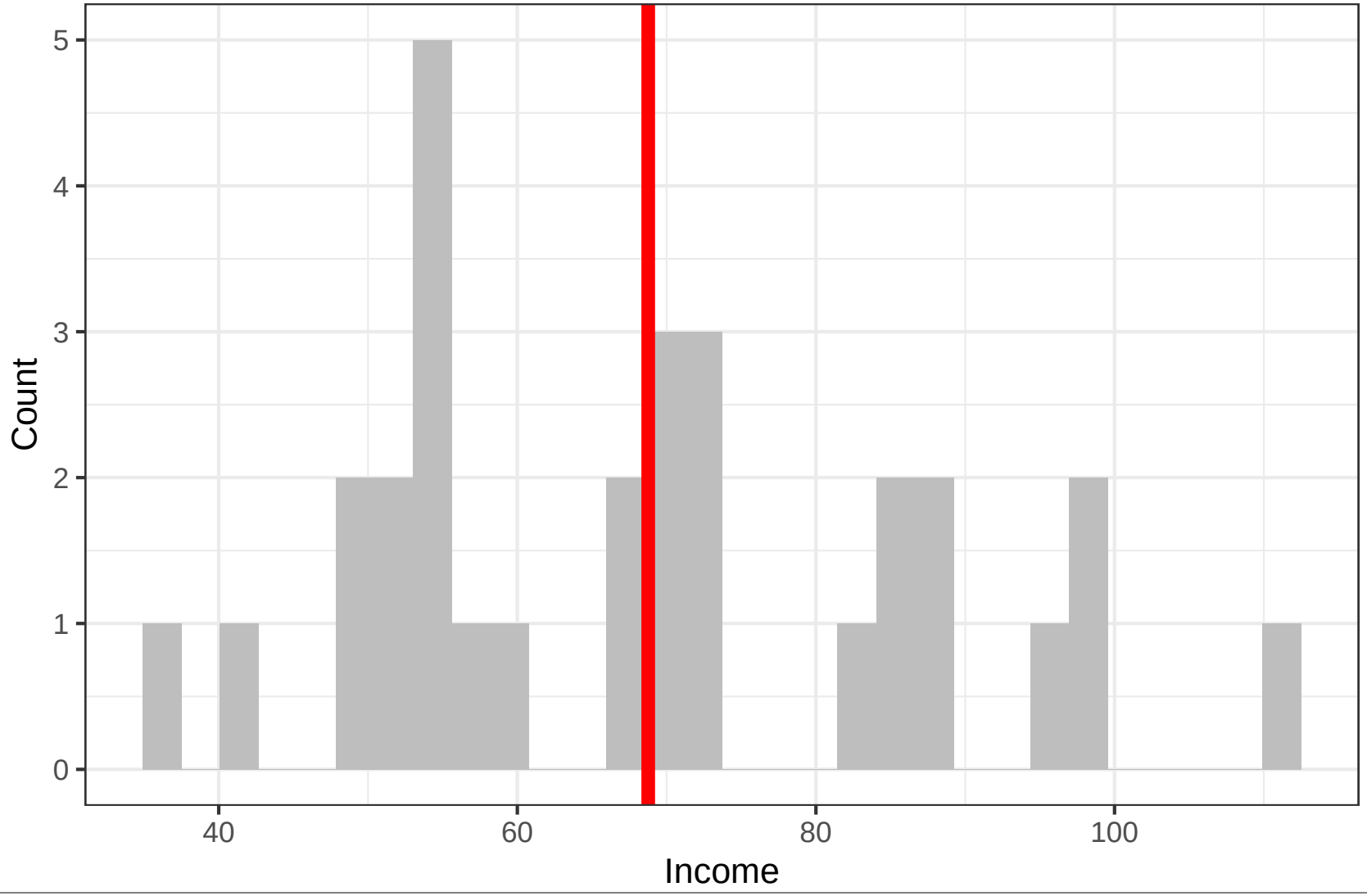
Statistical inference

- **Statistical inference** = a set of operations on data that yield estimates and uncertainty statements about **predictions and parameters** of some underlying process or population.
- More simply, we observe data and we try to learn something about where it came from.
- **This involves uncertainty.**

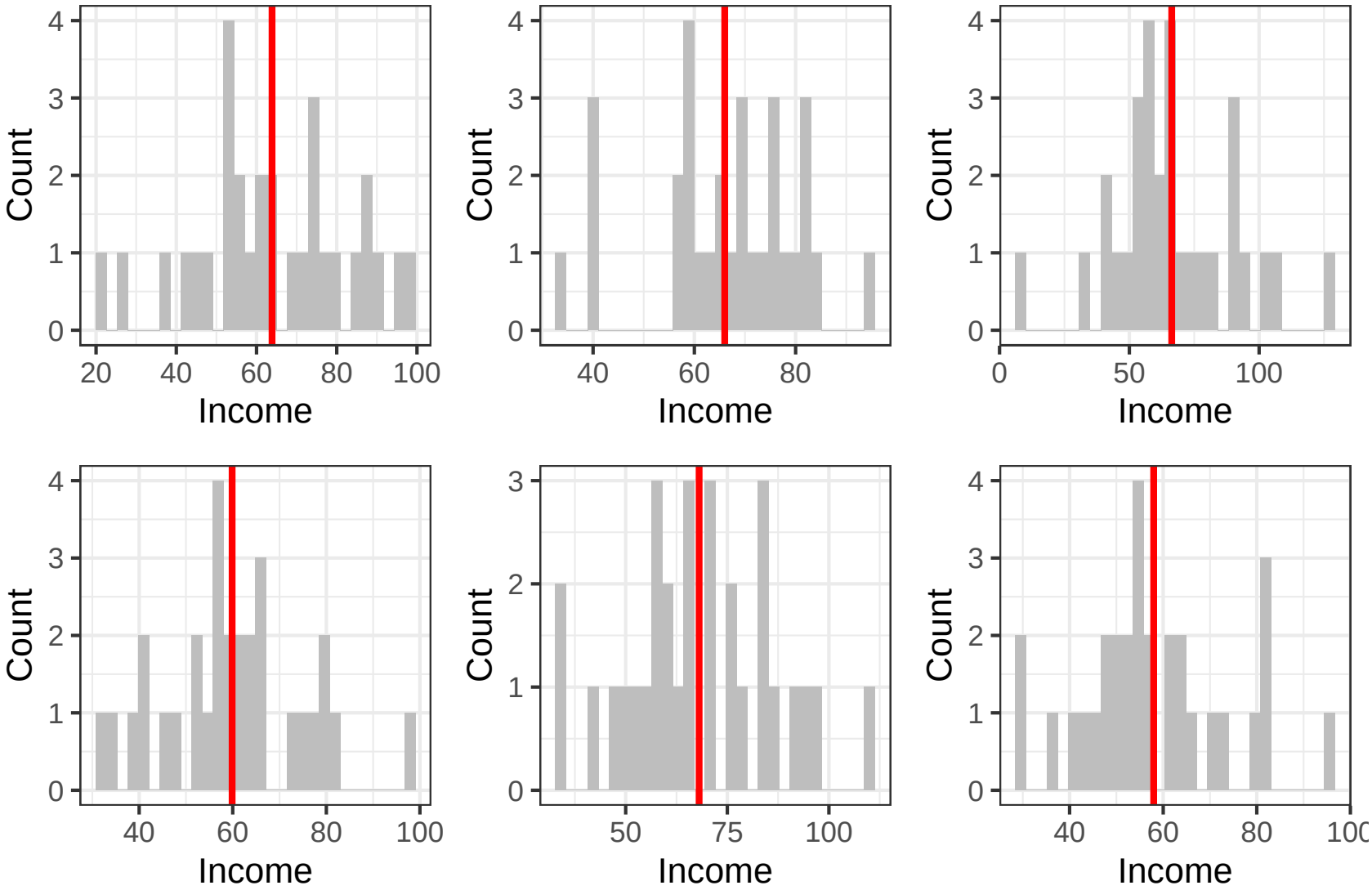
Statistical inference

- Say I have a population of 1000 people and I ask 30 of them their income.
- The parameter I'm interested in inferring something about is the average income in the entire population.

Statistical inference



Statistical inference

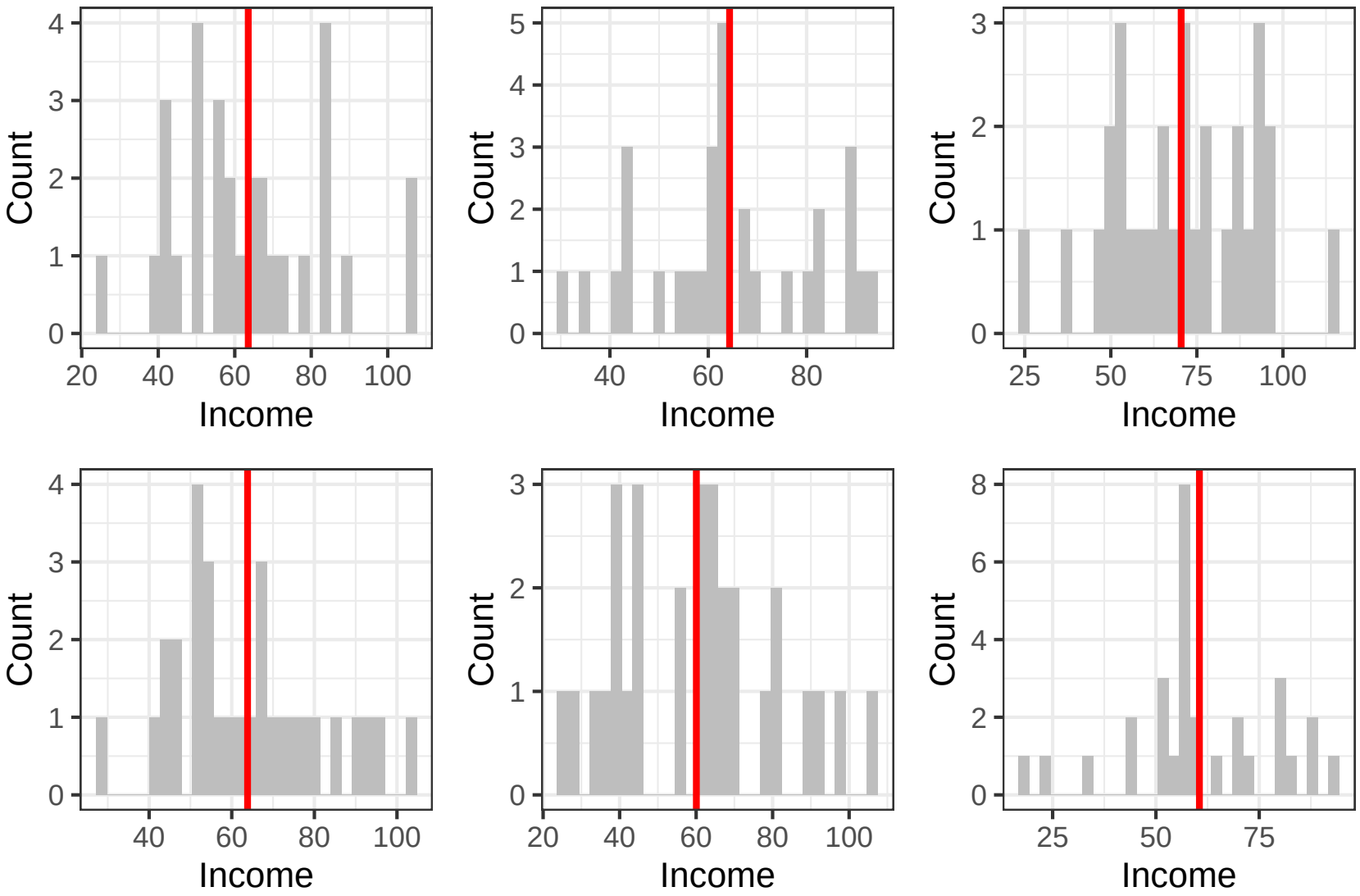


Sampling distribution

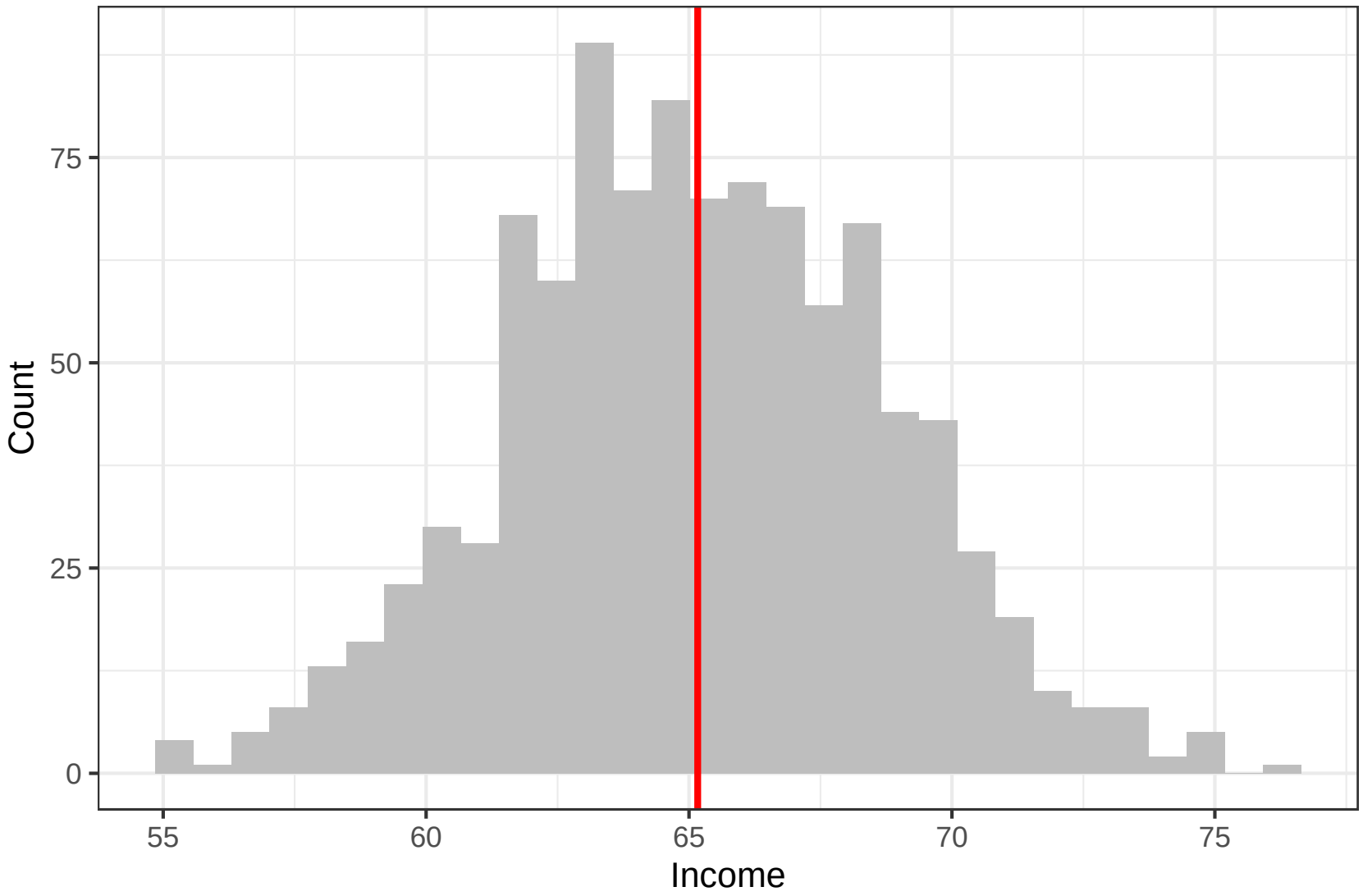
- **Sampling distribution** = the set of possible datasets that could have been observed if the data collection process had been re-done, along with the probabilities of these values.
- In practice, we will not know the sampling distribution; we can only estimate it.

Standard errors

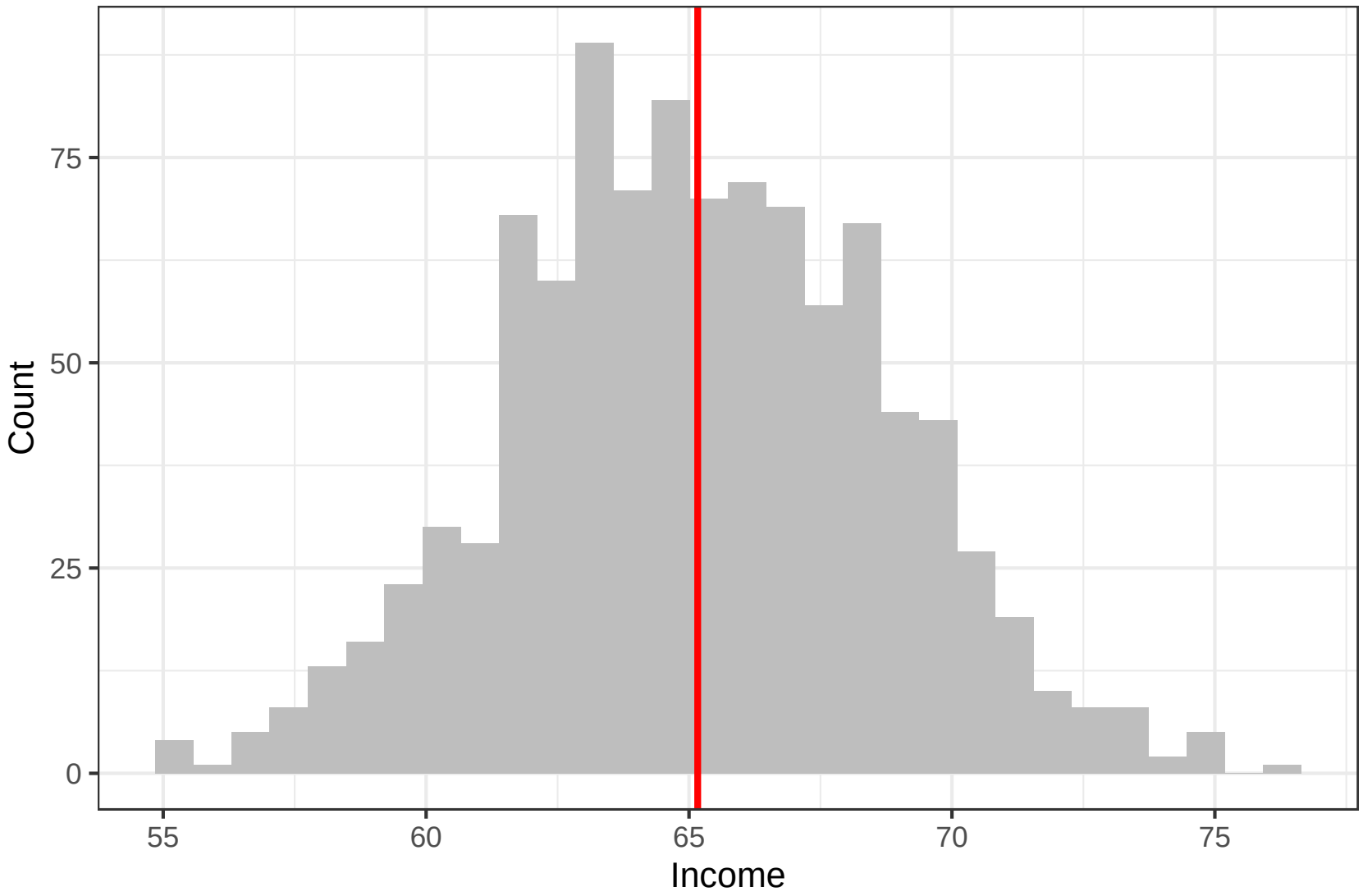
- **Standard error** = the estimated standard deviation of an estimate (e.g., the standard deviation of the sampling distribution).


Count
Income
0
1
2
3
4
5
6
8
20
25
40
50
60
75
80
100

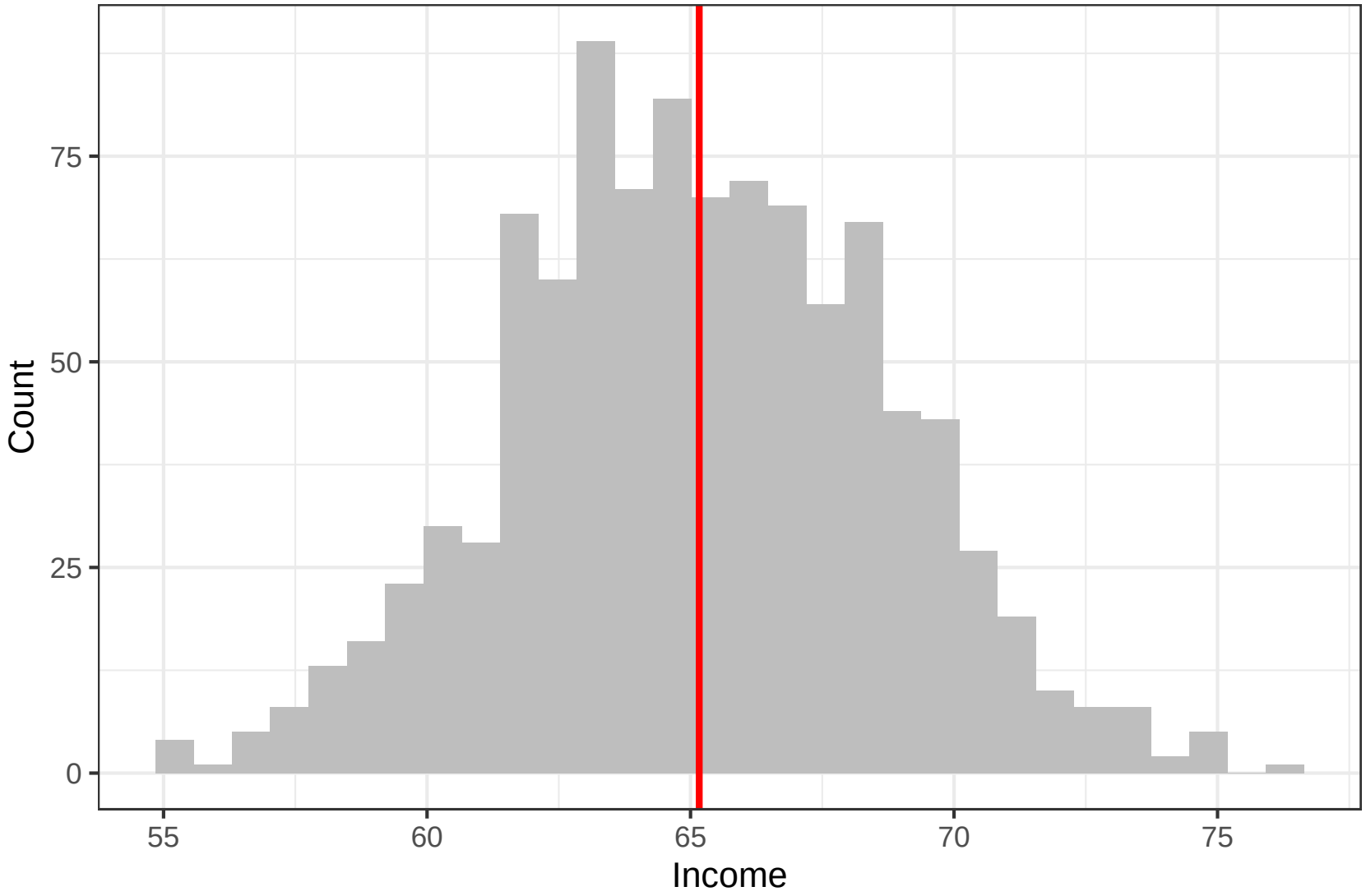
Sampling distribution of the mean:



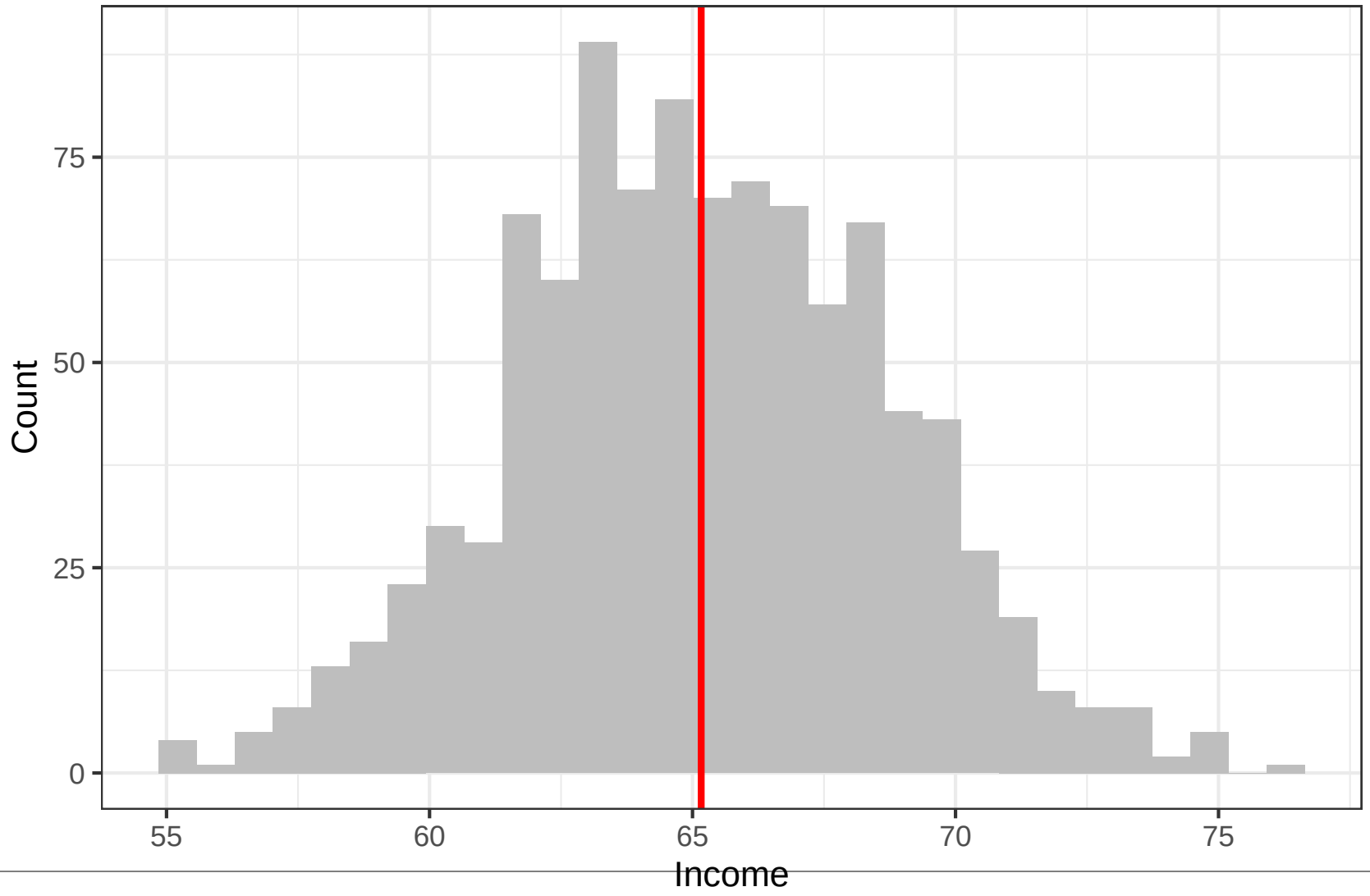
What do you notice about the shape of this distribution?



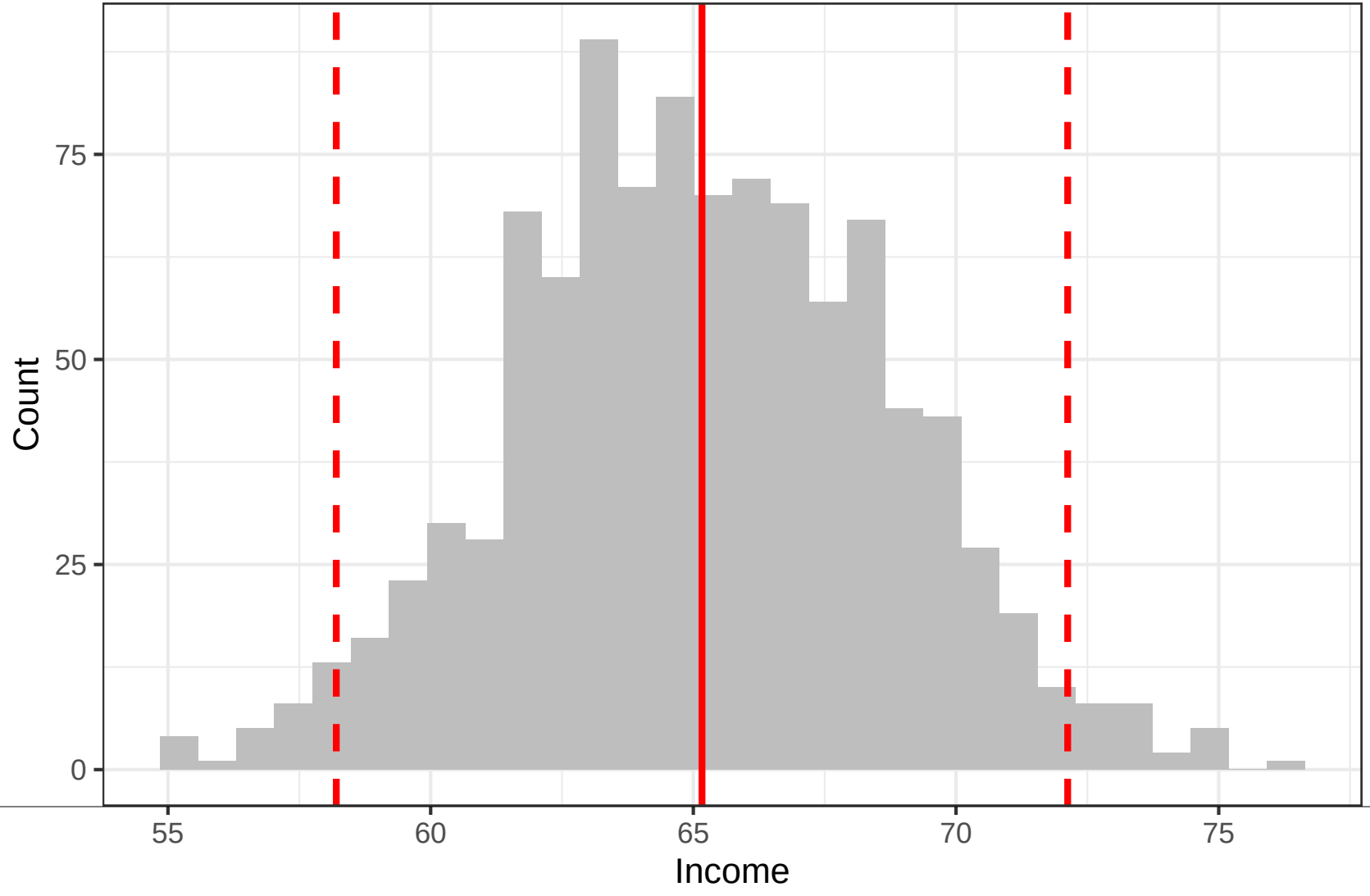
Central Limit Theorem



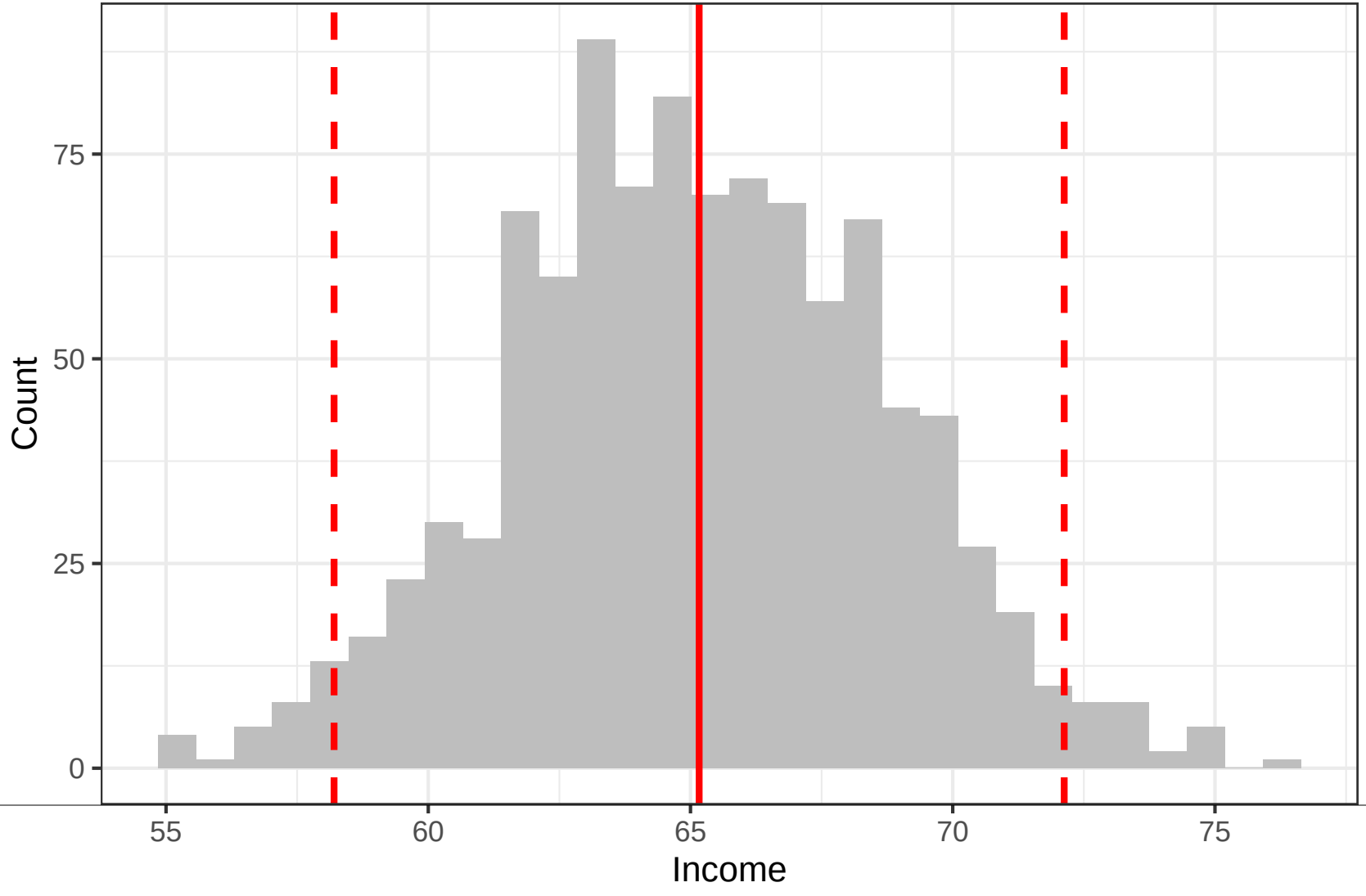
The **standard error** is the standard deviation of the sampling distribution of our target statistic.

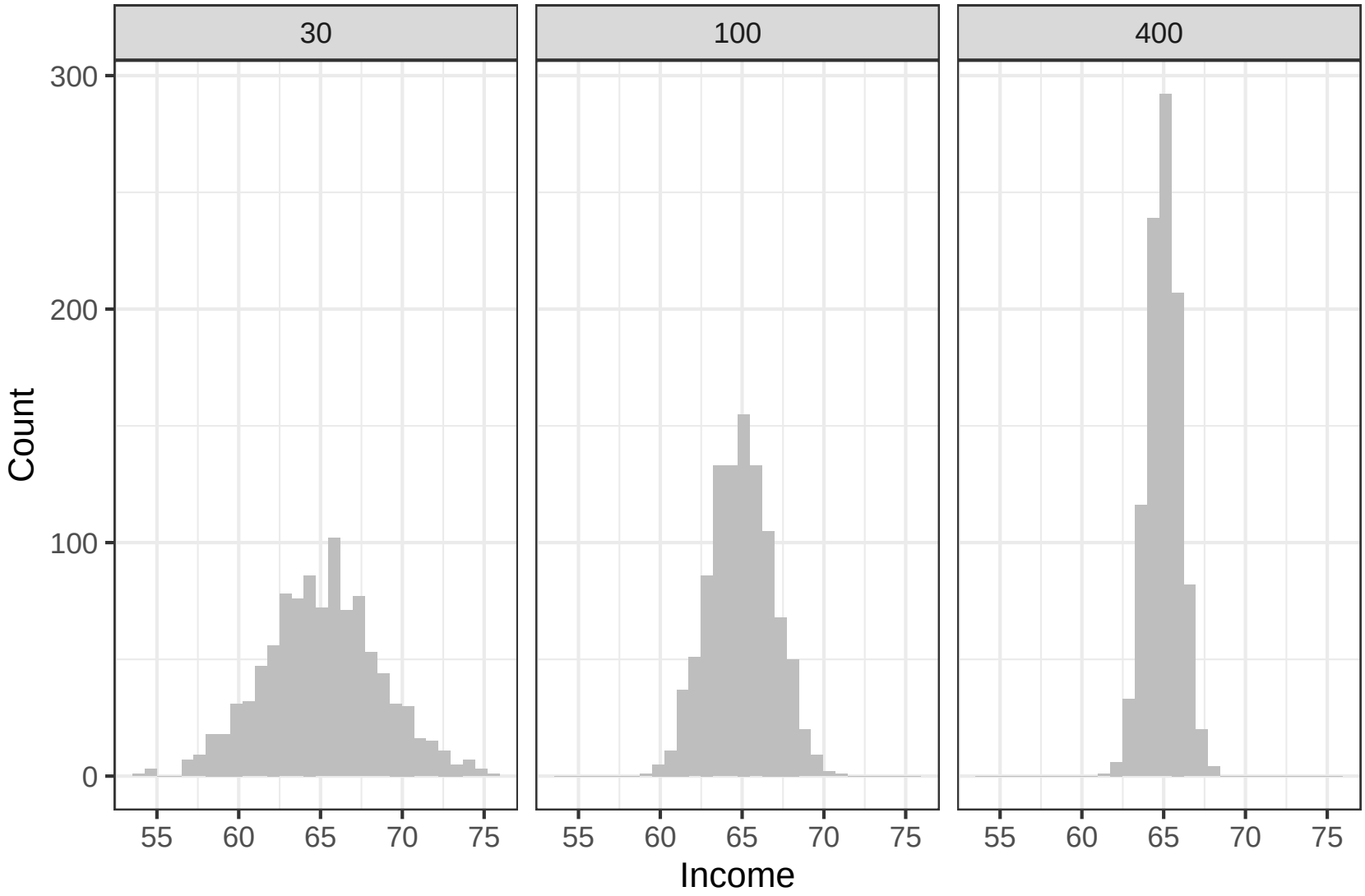


The confidence interval represents a range of values of a parameter or quantity of interest that are roughly consistent with the data, given the assumed sampling distribution.



However defined, the standard error is a measure of the variation in an estimate and gets smaller as sample size gets larger, converging on zero as the sample increases in size.




30
100
400
300
200
100
0
Count
55
60
65
70
75
Income