

# Simple Ain't Easy: Real-World Problems with Basic Summary Statistics

John Myles White and other contributors

---

## Problem 1

*The sampling distribution of the median can be multimodal if the source distribution is multimodal.*

## Example:

Let  $D$  be a distribution defined as a 50/50 mixture of two normals. As a specific example, we will assume that  $D$  is a mixture of two normals:  $\mathcal{N}(-10, 1)$  and  $\mathcal{N}(+10, 1)$ . The PDF for this distribution is shown in Figure **median/001/001**:

For this bimodal source distribution,  $D$ , the sampling distribution is not close to being normally distributed, because the median is most likely to be defined by either two points from the first mixture component or by two points from the second mixture component. This can be seen in the simulation results shown below:

In this case, the contrast between the erratic behavior of the sample mean and the sample median is very stark. As predicted by the Central Limit Theorem, the sampling distribution of the mean is approximately normal:

In the table below, we show the estimated standard deviations of the sampling distributions of the sample median and sample mean for this case. The median is roughly 20x more variable:

---

Std. Dev. of Sample Median	Std. Dev. of Sample Mean
2.918197	0.1612805

---

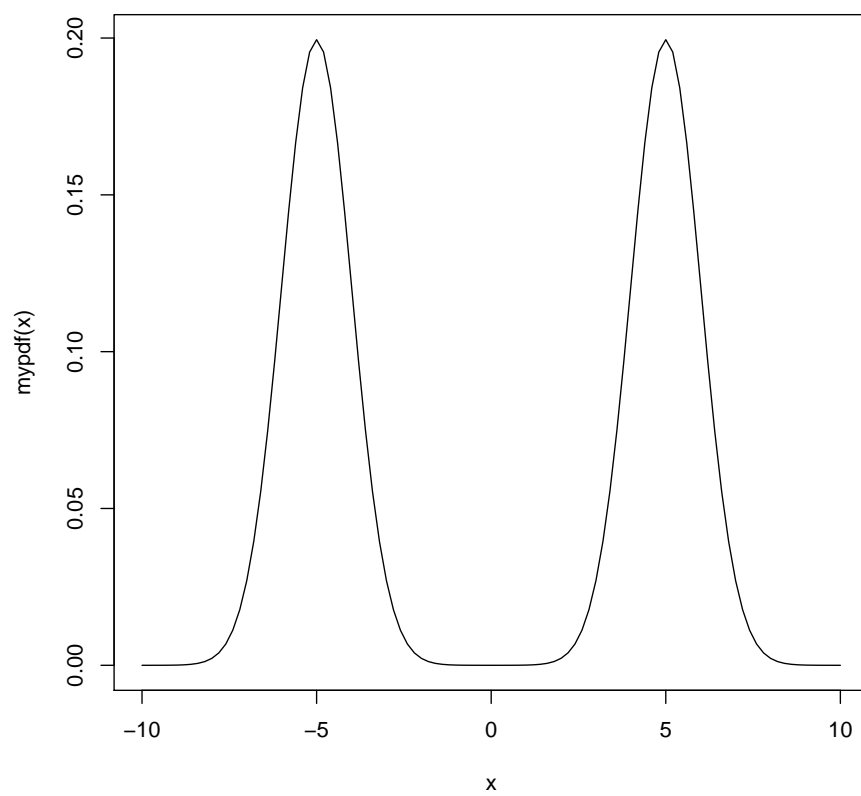


Figure 1: Probability Density Function of Source Distribution

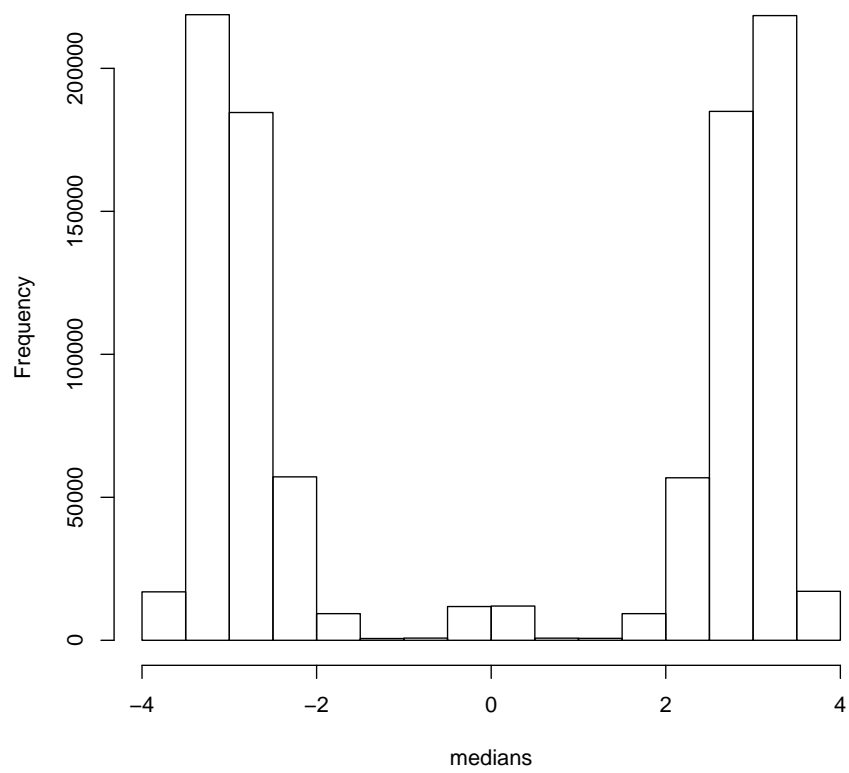


Figure 2: Histogram of the Sampling Distribution of the Median

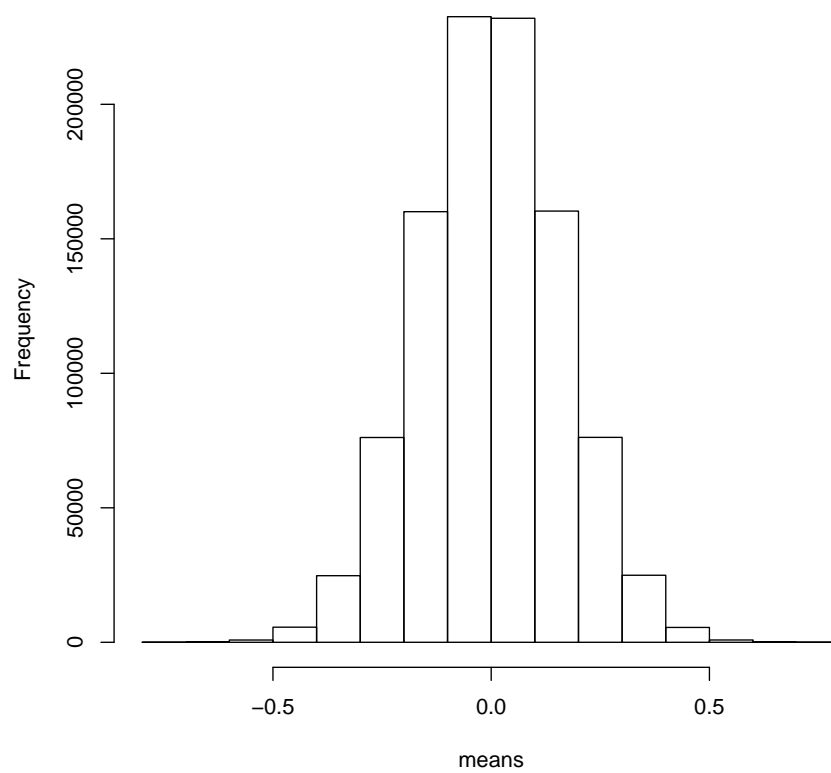


Figure 3: Histogram of the Sampling Distribution of the Mean