

Why divide by $n-1$? The sample variance mystery, solved

Lesson 80 · Module 6 — Estimation · Probability & Statistics · dailymaths.uk/probstat

Practice problems

1. Given the sample data $X = \{2, 4, 6, 8\}$, calculate the unbiased sample variance s^2 .
2. A researcher takes a sample of size $n = 10$ and finds the biased variance (dividing by n) to be 18. What is the unbiased sample variance s^2 ?
3. Explain why the bias of the sample variance decreases as the sample size n increases, and calculate the ratio of the unbiased estimator to the biased estimator for $n = 100$.

Solutions

1. Given the sample data $X = \{2, 4, 6, 8\}$, calculate the unbiased sample variance s^2 .

- Calculate the sample mean: $\bar{X} = (2 + 4 + 6 + 8)/4 = 5$.
- Calculate squared deviations: $(2 - 5)^2 = 9$, $(4 - 5)^2 = 1$, $(6 - 5)^2 = 1$, $(8 - 5)^2 = 9$.
- Sum the squared deviations: $9 + 1 + 1 + 9 = 20$.
- Divide by $n - 1 = 3$: $s^2 = 20/3 \approx 6.67$.

Answer: 6.67

2. A researcher takes a sample of size $n = 10$ and finds the biased variance (dividing by n) to be 18. What is the unbiased sample variance s^2 ?

- The biased variance is $s_{\text{biased}}^2 = 18$.
- The relationship is $s^2 = \frac{n}{n-1} s_{\text{biased}}^2$.
- Substitute the values: $s^2 = \frac{10}{9} \times 18$.
- Calculate: $s^2 = 10 \times 2 = 20$.

Answer: 20

3. Explain why the bias of the sample variance decreases as the sample size n increases, and calculate the ratio of the unbiased estimator to the biased estimator for $n = 100$.

- The bias exists because the sample mean is used instead of the population mean, which underestimates the spread.
- The correction factor is $\frac{n}{n-1}$.
- As $n \rightarrow \infty$, the term $\frac{n}{n-1} \rightarrow 1$, meaning the correction becomes smaller.
- For $n = 100$, the ratio is $100/99 \approx 1.0101$.

Answer: 1.0101