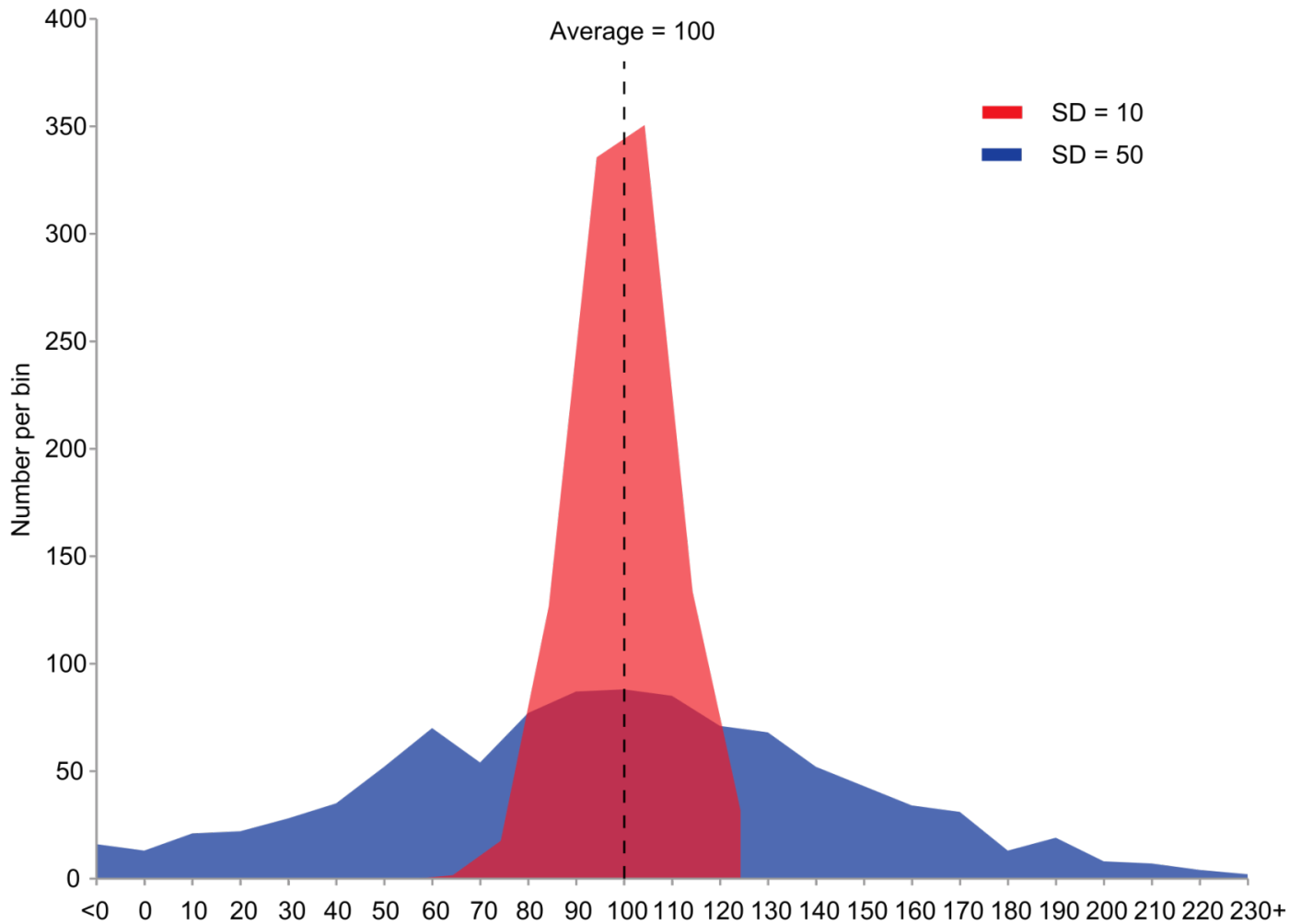


Variance

Variance essentially refers to how spread out your data is relative to its mean.



In the diagram the red distribution has low variance and the blue distribution has high variance.

In finance a probability distribution with high variance is typically seen as more risky (historical events have been very widely spread which implies that there is more chance that future events are more likely to fall across a much wider range).

Effects of Variance on ML Training Data

Data with high variance leads to less sensitive models and vice versa. This is really nicely illustrated in [this blog post](#) ([mirror](#))

Effects of Variance on Machine Learning Models

Models with a high degree of variance (often strongly tied to the number of parameters in the model) often fit better to training data but struggle to generalise to test data and vice versa.

Revision #1

Created 28 November 2022 11:13:11 by James

Updated 21 January 2024 14:49:27 by James