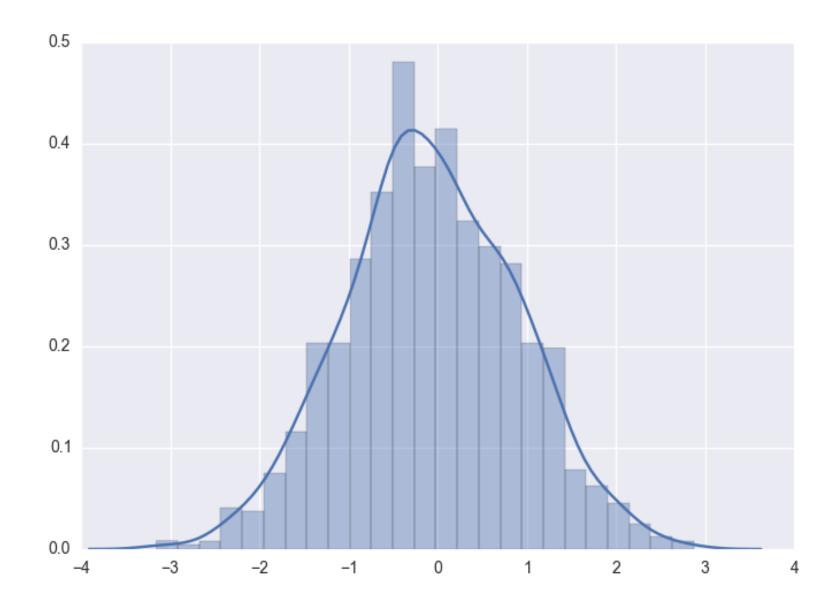# A statistical problem

1分

One common calculation in statistics is the variance (https://en.wikipedia.org/wiki/Variance) which is defined as the *mean of square minus square of mean*. Suppose we have a sample $x_0, x_1, \ldots, x_{n-1}$ of size n. Then the variance would be

$$\sigma^2 = \frac{1}{n}\left(\sum_{i=0}^{n} x_i^2 - \frac{(\sum_{i=0}^{n} x_i)^2}{n} \right)$$

Using `x = np.random.randn(1000)` a binned plot using `plt.hist(x, 25)` is shown in the following

How do we calculuate the variance above? Read the following section

- naive calculation of the variance
  (https://en.wikipedia.org/wiki/Algorithms_for_calculating_variance#Na.C3.AFve_algorithm)

Then read the following example

- an example of calculating the variance
  (https://en.wikipedia.org/wiki/Algorithms_for_calculating_variance#Example)

What is the issue with a basic approach to summation?

选项*

○ The algorithm leads to division-by-zero.

○ Cancellation of two similar numbers in a subtraction leads to loss-of-significance in the result.

○ The naive algorithm is simply too expensive.

○ Summation of large numbers leads to overflow, resulting in program failure.

参考答案: 'Cancellation of two similar numbers in a subtraction leads to loss-of-significance in the result.'.