Course > Probability: Discrete Random Variables > Mean and Variance > Variance and Standard Deviation of a Random Variable

🔖 Bookmark this page

Variance and Standard Deviation of a Random Variable

> **Learning Objective: Find the mean and variance of a discrete random variable, and apply these concepts to solve real-world problems.**

Variance and Standard Deviation of a Discrete Random Variable

In Exploratory Data Analysis, we used the mean of a sample of quantitative values (their arithmetic average, $\overline{x}$) to tell the center of their distribution, and the standard deviation (s) to tell the typical distance of sample values from their mean. We described the center of a probability distribution for a random variable by reporting its mean μ_X, and now we would like to establish an accompanying measure of spread. Our measure of spread will still report the typical distance of values from their means, but in order to distinguish the spread of a population of all of a random variable's values from the spread (s) of sample values, we will denote the standard deviation of the random variable X with the Greek lower case "σ," and use a subscript to remind us what is the variable of interest (there may be more than one in later problems):

Notation: σ_X

We will also focus more frequently than before on the squared standard deviation, called the **variance**, because some important rules we need to invoke are in terms of variance σ_X^2 rather than standard deviation σ_X.

Example: Xavier's Production Line

Recall that the number of defective parts produced each hour by Xavier's production line is a random variable X with the following probability distribution:

X	0	1	2	3	4
P(X=x)	.15	.30	.25	.20	.10

We found the mean number of defective parts produced per hour to be μ_X = 1.8. Obviously, there is variation about this mean: some hours as few as 0 defective parts are produced, whereas in other hours as many as 4 are produced. Typically, how far does the number of defective parts fall from the mean of 1.8? As we did for the spread of sample values, we measure the spread of a random variable by calculating the square root of the average squared deviation from the mean. Now "average" is a weighted average, where more probable values of the random variable are accordingly given more weight. Let's begin with the variance, or average squared deviation from the mean, and then take its square root to find the standard deviation:

Values of X	0	1	2	3	4
Dev. from mean	(0-1.8)	(1-1.8)	(2-1.8)	(3-1.8)	(4-1.8)
Sq. deviations	$(0-1.8)^2$	$(1-1.8)^2$	$(2-1.8)^2$	$(3-1.8)^2$	$(4-1.8)^2$
Probabilities	.15	.30	.25	.20	.10

$$\text{Variance} = \sigma_X^2 = (0 - 1.8)^2(0.15) + (1 - 1.8)^2(0.30) + (2 - 1.8)^2(0.25) + (3 - 1.8)^2(0.20) + (4 - 1.8)^2(0.1) = 1.46$$

$$\text{standard deviation} = \sigma_X = \sqrt{1.46} = 1.21$$

How do we interpret the standard deviation of X?

Xavier's production line produces an average of 1.80 defective parts per hour. The number of defective parts varies from hour to hour; typically (or, on average), it is about 1.21 away from 1.80.

Here is the formal definition:

standard deviation of a discrete random variable

(definition) For any discrete random variable X with a probability distribution of

X	x_1	x_2	x_3	$\cdots$	x_n
P(X=x)	p_1	p_2	p_3	$\cdots$	p_n

the **variance** of X is defined to be $\sigma_X^2 = (x_1 - \mu_X)^2 p_1 + (x_2 - \mu_X)^2 p_2 + \ldots + (x_n - \mu_X)^2 p_n = \sum_{i=1}^{n} (x_i - \mu_X)^2 p_i$ and the **standard deviation** is $\sigma_X = \sqrt{\sigma_X^2}$


The purpose of the next activity is to give you better intuition about the mean and standard deviation of a random variable.

Keeping in mind that the mean describes where a histogram is centered, and the standard deviation describes spread by reporting the typical distance of values from their mean, compare the histograms in the four exercises below and match each to the correct combination of mean and standard deviation.

Learn By Doing

1/1 point (graded)
What is the mean and standard deviation of this histogram?
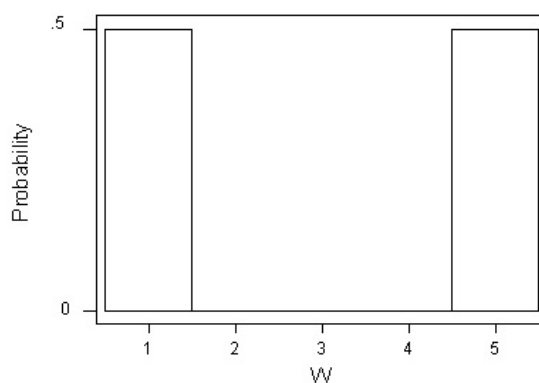
- ○ mean 3 and standard deviation 1
- ● mean 3 and standard deviation 2 ✔

○ mean 4 and standard deviation 1
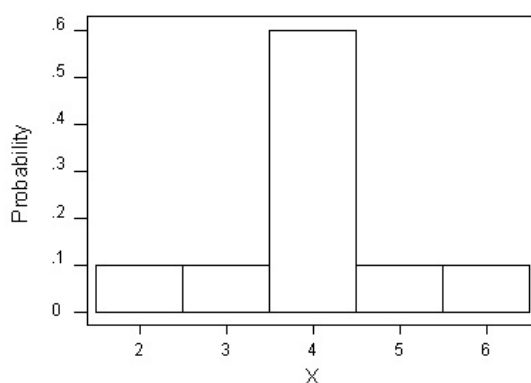
○ mean 4 and standard deviation 2

Answer
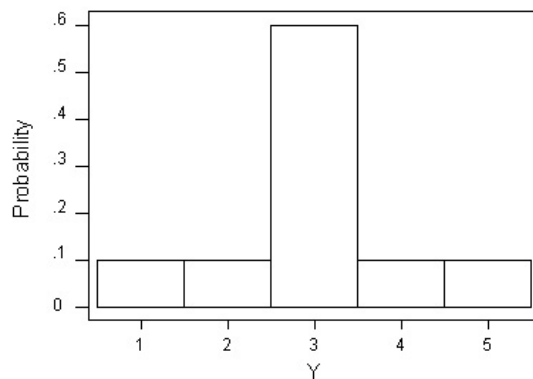Correct:
Indeed, the probability distribution of the random variable W is centered at the mean of 3, and the typical distance between the values that W can have (1,5) and the mean 3 is 2.

Submit

Learn By Doing

1/1 point (graded)
What is the mean and standard deviation of this histogram?

○ mean 3 and standard deviation 1

○ mean 3 and standard deviation 2

◉ mean 4 and standard deviation 1 ✔

○ mean 4 and standard deviation 2

Answer
Correct:
Indeed, the probability distribution of the random variable X is centered at the mean of 4, and the typical distance between the values that X can have and the mean is 1.

Submit

Learn By Doing

1/1 point (graded)
What is the mean and standard deviation of this histogram?

- ⦿ mean 3 and standard deviation 1 ✔
- ○ mean 3 and standard deviation 2
- ○ mean 4 and standard deviation 1
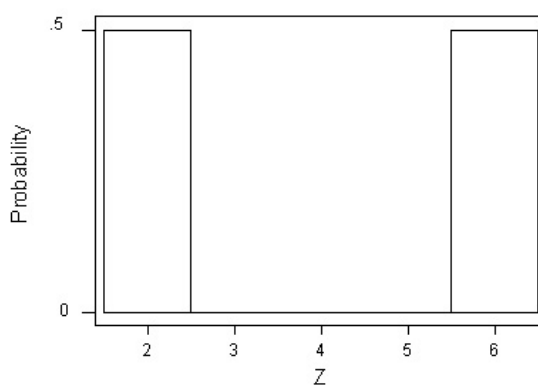- ○ mean 4 and standard deviation 2

Answer
Correct:
Indeed, the probability distribution of the random variable Y is centered at the mean of 3, and the typical distance between the values that Y can have and the mean is 1.

Submit

Learn By Doing

1/1 point (graded)
What is the mean and standard deviation of this histogram?

- ○ mean 3 and standard deviation 1
- ○ mean 3 and standard deviation 2

⚪ mean 4 and standard deviation 1

🔘 mean 4 and standard deviation 2 ✔

Answer
Correct:
Indeed, the probability distribution of the random variable Z is centered at the mean of 4, and the typical distance between the values that Z can have and the mean is 2.

Submit

Open Learning Initiative ⧉