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

🔖 Bookmark this page

Mean and Variance of a Random Variable: Introduction

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

In the Exploratory Data Analysis (EDA) section, we displayed the distribution of one quantitative variable with a histogram, and supplemented it with numerical measures of center and spread. We are doing the same thing here. We display the probability distribution of a discrete random variable with a table, formula or histogram, and supplement it with numerical measures of the center and spread of the probability distribution. These measures are the **mean and standard deviation of the random variable.**

This section will be devoted to introducing these measures. As before, we'll start with the numerical measure of center, the mean. Let's begin by revisiting an example we saw in EDA.

World Cup Soccer

Recall that we used the following data from 3 World Cup tournaments (a total of 192 games) to introduce the idea of a **weighted average**.

We've added a third column to our table that gives us relative frequencies.

total # goals/game	frequency	relative frequency
0	17	17 / 192 = 0.089
1	45	45 / 192 = 0.234
2	51	51 / 192 = 0.266
3	37	37 / 192 = 0.193

4	25	25 / 192 = 0.130
5	11	11 / 192 = 0.057
6	3	3 / 192 = 0.016
7	2	2 / 192 = 0.010
8	1	1 / 192 = 0.005

the mean for this data is:

$$\frac{0(17)+1(45)+2(51)+3(37)+4(25)+5(11)+6(3)+7(2)+8(1)}{192}$$

distributing the division by 192 we get:

$$0\left(\frac{17}{192}\right) + 1\left(\frac{45}{192}\right) + 2\left(\frac{51}{192}\right) + 3\left(\frac{37}{192}\right) + 4\left(\frac{25}{192}\right) + 5\left(\frac{11}{192}\right) + 6\left(\frac{3}{192}\right) + 7\left(\frac{2}{192}\right) + 8\left(\frac{1}{192}\right)$$

Notice that the mean is each number of goals per game multiplied by its relative frequency. Since we usually write the relative frequencies as decimals, we can see that:

mean number of goals per game = 0(0.089) + 1(0.234) + 2(0.266) + 3(0.193) + 4(0.130) + 5(0.057) + 6(0.016) + 7(0.010) + 8(0.005)

mean number of goals per game = 2.36, rounded to two decimal places

Mean of a Random Variable

In Exploratory Data Analysis, we used the mean of a sample of quantitative values—their arithmetic average—to tell the center of their distribution. We also saw how a weighted mean was used when we had a frequency table. These frequencies can be changed to relative frequencies. So we are essentially using the relative frequency approach to find probabilities. We can use this to find the mean, or center, of a probability distribution for a random variable by reporting its mean, which will be a weighted average of its values; the more probable a value is, the more weight it gets. As always, it is important to distinguish between a concrete sample of observed values for a variable versus an abstract population of all values taken by a random variable in the long run.

Whereas we denoted the mean of a sample as $\overline{x}$, we now denote the mean of a random variable as μ_X. Let's see how this is done by looking at a specific example.

> ### Example: Xavier's Production Line

Xavier's production line produces a variable number of defective parts in an hour, with probabilities shown in this table:

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

How many defective parts are typically produced in an hour on Xavier's production line? If we sum up the possible values of X, each weighted with its probability, we have

$$\mu_X = 0(0.15) + 1(0.30) + 2(0.25) + 3(0.20) + 4(0.10) = 1.8$$

Here is the general definition of the mean of a discrete random variable:

mean of a discrete random variable

(definition) In general, for any discrete random variable X with probability distribution

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

the **mean** of X is defined to be $\mu_X = x_1 p_1 + x_2 p_2 + \ldots + x_n p_n = \sum_{i=1}^{n} x_i p_i$

In general, the mean of a random variable tells us its "long-run" average value. It is sometimes referred to as the **expected value** of the random variable. But this expression may be somewhat misleading, because in many cases it is impossible for a random variable to actually equal its expected value. For example, the mean number of goals for a World Cup soccer game is 2.36. But we can never expect any single game to result in 2.36 goals, since it is not possible to score a fraction of a goal. Rather, 2.36 is the long-run average of all World Cup soccer games. In the case of Xavier's production line, the mean number of defective parts produced in an hour is 1.8. But the actual number of defective parts produced in any given hour can never equal 1.8, since it must take whole number values.

To get a better feel for the mean of a random variable, let's extend the defective parts example:

Example: Xavier's and Yves' Production Lines

Recall the probability distribution of the random variable X, representing the number of defective parts in an hour produced by Xavier's production line.

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

The number of defective parts produced each hour by Yves' production line is a random variable Y with the following probability distribution:

Y	0	1	2	3	4
P(Y=y)	.05	.05	.10	.75	.05

Look at both probability distributions. Both X and Y take the same possible values (0, 1, 2, 3, 4).

However, they are very different in the way the probability is distributed among these values.

Learn By Doing

1/1 point (graded)

Recall that the mean number of defective products in an hour for Xavier's production line (the mean of X) was 1.8. Based on your observation, which of the following do you think is true about the mean of Y?

- ●   The mean of Y is greater than 1.8 ✔

- ○   The mean of Y is less than 1.8

- ○   The mean of Y is roughly equal to 1.8

Answer
Correct:
Indeed, the probability distribution of Y has most of the weight (80%) on the larger values (3, 4), while the distribution of X has most of the weight (70%) on the smaller values (0, 1, 2). Since the mean is an average of the values weighted by their probabilities, we would expect the mean of Y to be larger than the mean of X, which is 1.8.

Submit

Did I Get This

1/1 point (graded)

Here again is the probability distribution of Y, the number of defective parts in an hour in Yves' production line:

Y	0	1	2	3	4
P(Y=y)	.05	.05	.10	.75	.05

What is the mean of Y?

○ 2

○ 0.2

● 2.7 ✔

○ 0.54

Answer

Correct: Indeed, the mean of Y is 0(.05) + 1(.05) + 2(.10) + 3(.75) + 4(.05) = 2.7.

Submit

Open Learning Initiative ⬈