6.19: Discrete Random Variables (4 of 5)

Learning Objectives

- Use probability distributions for discrete and continuous random variables to estimate probabilities and identify unusual events.

Standard Deviation for a Discrete Random Variable

The mean of a discrete random variable gives us a measure of the long-run average but it gives us no information at all about how much variability to expect. For example, earlier we found that the average cafeteria wait time at Rushmore Community College was 14 minutes. Put in terms of our random variable, this means over the long run, if we continued to keep track of wait times for students entering the cafeteria, their times would average 14 minutes. Some students would get their food in less than 14 minutes, and some would have to wait longer.

Is that all we need to know? Suppose on the one hand the average time was 14 minutes, but we knew that it was most likely that times would range from 8 to 20 minutes. Compare that to a situation where again the average time was 14 minutes, but it was most likely that times would range only from 13 to 15 minutes. That would give us a different picture of what the problem at the cafeteria might be. What we need is a measure of how much variability to expect in a random variable *X* over the long run. The standard deviation is that measure.

Just as we need both the mean and standard deviation to get a full picture of the shape of a data set, we need both the mean and standard deviation of a random variable to understand its likely long-term behavior.

In *Summarizing Data Graphically and Numerically*, we used the following formula to compute the standard deviation of a data set.

$$\sqrt{\frac{\sum(x-\bar{x})^{2}}{n-1}}$$

As you may recall, the most important part of this formula is the term inside the square root, which we call the *average of the squares of the deviations from the mean*.

$$\frac{\sum(x-\bar{x})^{2}}{n-1}$$

As we will see, the formula for the standard deviation for a discrete random variable has a lot in common with this formula.

Here is the formula for the standard deviation of a discrete random variable. Note that $P(x)$ represents the probability of *x*, where *x* is a value of the random variable *X*. And ${\mu}_{x}$ again stands for the mean of *X*.

$$\sqrt{\Sigma(x-{\mu}_{x})^{2}p(x)}$$

Again, we focus on the term inside the square root:

$$\Sigma(x-{\mu}_{x})^{2}p(x)$$

The term $(x-{\mu}_{x})$ here represents the deviation of each value of the random variable *X* from the mean ${\mu}_{x}$, just as the term $(x-\bar{x})$ represents the deviation of each observation of the data set from the mean $\bar{x}$.

In both cases, we proceed to sum the squares of these deviations. In the case of a data set, we divide by $n-1$ to find the average squared deviation. However, in the case of a discrete random variable, we again use a weighted average. Why? Because we don't want to give undue weight to values of *X* that are unlikely to occur. So those values of *X*, even if far from the mean ${\mu}_{x}$, will not contribute much to the standard deviation if their probability is low. On the other hand, values of *X* with large probabilities will count more in our calculation of the standard deviation of *X*.

Example

Cafeteria Wait Times

Let's revisit the problem about wait times in the cafeteria at Rushmore Community College. Recall the following probability distribution.

X = Time (minutes)	5	10	15	20	25
P(*X*)	**0.15**	**0.26**	**0.31**	**0.20**	**0.08**

On the previous page, we found that the average wait time is 14 minutes. Now we will compute the standard deviation of wait times and think a bit about what it tells us.

We start by computing the squared deviations from the mean and weighting them by the probability. For the first value of X, we have

$${(x-{\mathrm{\mu}}_{x})}^{2}\cdot p(x)={(5-14)}^{2}\cdot(0.15)={(-9)}^{2}\cdot(0.15)=81\cdot(0.15)=12.15$$

Performing the same operation on the next three values of X will give us

$$\begin{array}{l}{(10-14)}^{2}\cdot(0.26)=4.16\\ {(15-14)}^{2}\cdot(0.31)=0.31\\ {(20-14)}^{2}\cdot(0.20)=7.2\end{array}$$

Try It

https://assessments.lumenlearning.co...sessments/3560

The next step of the formula is to add up the weighted square deviations from the mean, as follows:

Try It

https://assessments.lumenlearning.co...sessments/3561

Recall that in *Summarizing Data Graphically and Numerically* we used the standard deviation of a quantitative data set to give a range of typical values. This range of typical values was formed by blocking off an interval 1 standard deviation to the right and left of the mean. In other words, the range of typical values was $\left[\bar{x}-1\cdot\mathrm{SD},\bar{x}+1\cdot\mathrm{SD}\right]$. Exactly the same thing can be done in the current context of random variables.

Try It

https://assessments.lumenlearning.co...sessments/3562

Contributors and Attributions