Lesson 29: What is a probability model?

Define a variable for each scenario and think about the possbile outcomes?

1. A grocery store manager might be interested in the number of broken eggs in each carton (dozen of eggs).

   OR

2. An environmental scientist might be interested in the amount of ozone in an air sample.

Random variable

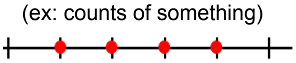
A numerical variable whose value depends on the outcome of a chance experiment

2 types

Discrete

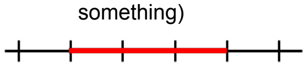
When the possible values can be listed

(ex: counts of something)



Continuous

When the set of possible values includes and entire interval (measure of something)



Identify the following variables as discrete or continuous

1. The number of broken eggs in each carton
Discrete
2. The amount of ozone in samples of air
Continuous
3. The weight of a pineapple
Continuous
4. The amount of time a customer spends in a store
Continuous
5. The number of gas pumps in use
Discrete

Probability Distributions for Discrete Random Variables

Probability distribution is a model that describes the long-run behavior of a variable.

The collection of all the possible values and the probabilities that they occur

In a Wolf City (a fictional place), regulations prohibit no more than five dogs or cats per household.
Let **x** = the number of dogs and cats in a randomly selected household in Wolf City

Is the random variable discrete or continuous?

What are the possible values of x?

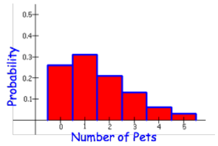
x	0	1	2	3	4	5
P(x)	.26	.31	.21	.13	.06	.03

The Department of Animal Control has collected data over the course of several years. They have estimated **the long-run probabilities** for the values of x.

In a Wolf City (a fictional place), regulations prohibit no more than five dogs or cats per household.

x	0	1	2	3	4	5
P(x)	.26	.31	.21	.13	.06	.03

What's the sum of p(x)?



1) For every possible x value, $0 \le P(x) \le 1$.
2) For all values of x, $\Sigma P(x) = 1$.

Dogs and Cats Revisited . . .

What is the probability that a randomly selected household in Wolf City has at most 2 pets?

x	0	1	2	3	4	5
P(x)	.26	.31	.21	.13	.06	.03

$P(x \le 2) = .26 + .31 + .21 = .78$

What is the probability that a randomly selected household in Wolf City has less than 2 pets?

$P(x < 2) = .26 + .31 = .57$

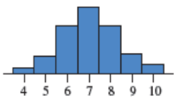
What is the probability that a randomly selected household in Wolf City has more than 1 but no more than 4 pets?

$P(1 < x \le 4) = .21 + .13 + .06 = .40$

Probability Distributions for Continuous Random Variables

Consider the random variable:

x = the weight (in pounds) of a full-term newborn child



Suppose that weight is reported to the nearest pound.

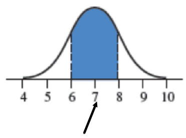
The area of the rectangle centered over 7 pounds represents the probability $6.5 \le x \le 7.5$

Now suppose that weight is reported to the nearest 0.1 pound. This would be the probability histogram.

Greater accuracy

-> narrower rectangles

->smoother curve



Density curve : specifies prob. distro. for continuous vars.

The shaded area represents the probability $6 \le x \le 8$.

What is the sum of the areas of all the rectangles?

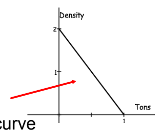
or under density curve? 1

Let x denote the amount of gravel sold (in tons) during a randomly selected week at a particular sales facility. Suppose that the density curve has a height f(x) above the value x, where

$$f(x) = \begin{cases} 2(1-x) & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

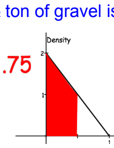
Density Function

Density curve



What is the probability that at most ½ ton of gravel is sold during a randomly selected week?

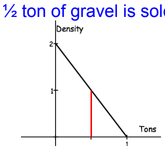
$P(x \le \frac{1}{2}) =$ $1 - \frac{1}{2}(0.5)(1) = .75$



*total area under density curve is 1

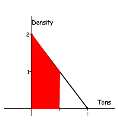
What is the probability that exactly ½ ton of gravel is sold during a randomly selected week?

$P(x = \frac{1}{2}) =$



What is the probability that less than ½ ton of gravel is sold during a randomly selected week?

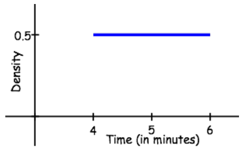
$P(x < \frac{1}{2}) = P(x \le \frac{1}{2})$



Remember this was different in discrete!

Suppose x is a continuous random variable defined as the amount of time (in minutes) taken by a clerk to process a certain type of application form. Suppose x has a probability distribution with density function:

$$f(x) = \begin{cases} .5 & 4 < x < 6 \\ 0 & \text{otherwise} \end{cases}$$



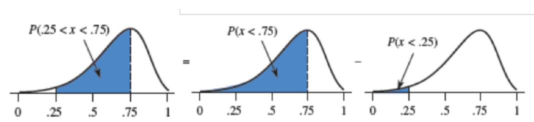
What is the probability that it takes more than 5.5 minutes to process the application form?

$P(x > 5.5) = .5(.5) = .25$

In general,

The probability that a continuous random variable x lies between a lower limit a and an upper limit b is

$$P(a < x < b) = P(x < b) - P(x < a)$$



Mean and Variance for Discrete Probability Distributions

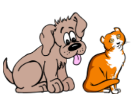
Mean is sometimes referred to as the expected value (denoted E(x)).

$$\mu_x = \sum xp$$

Variance is calculated using

$$\sigma^2 = \sum (x - \mu_x)^2 p$$

Standard deviation is the square root of the variance.



Let x = the number of dogs and cats in a randomly selected household in Wolf City

x	0	1	2	3	4	5
P(x)	.26	.31	.21	.13	.06	.03

What is the mean number of pets per household in WolfCity?

$$\mu_x = \sum xp$$

What is the standard deviation of the number of pets per household in Wolf City?

$$\sigma^2 = \sum (x - \mu_x)^2 p$$

$\sigma_x^2 =$

Find SD?

Bring a deck!

Pick a card, any card. You draw a card from a deck. If you get a red card, you win nothing. If you get a spade, you win $5. For any club, you win $10 plus an extra $20 for the ace of clubs.

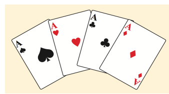
a) Create a probability model for the amount you win.
b) Find the expected amount you'll win.
c) What would you be willing to pay to play this game?

1. All the students stand up.
2. You start rolling a die.
3. Each student gets the number of points on the die. Students then may keep their points by sitting down, or remain standing for another roll.
4. With each roll, points accumulate for those who choose to remain standing. BUT . . . if you roll a 5, all those still standing lose their points, ending the round with a score of 0.

Who can amass the highest score?

Strategy?

On Valentine's Day the *Quiet Nook* restaurant offers a *Lucky Lovers Special* that could save couples money on their romantic dinners. When the waiter brings the check, he'll also bring the four aces from a deck of cards. He'll shuffle them and lay them out face down on the table. The couple will then get to turn one card over. If it's a black ace, they'll owe the full amount, but if it's the ace of hearts, the waiter will give them a $20 Lucky Lovers discount. If they first turn over the ace of diamonds (hey—at least it's red!), they'll then get to turn over one of the remaining cards, earning a $10 discount for finding the ace of hearts this time.



Build a probability model:

Outcome			
x			
$P(X = x)$			

Check results page 386

Means and Variances: Adding and Subtracting

Adding or subtracting a constant from data shifts the mean but doesn't change the variance or standard deviation.

$$E(X \pm c) = E(X) \pm c \qquad Var(X \pm c) = Var(X)$$

We've determined that couples dining at the *Quiet Nook* can expect Lucky Lovers discounts averaging $5.83 with a standard deviation of $8.62. Suppose that for several weeks the restaurant has also been distributing coupons worth $5 off any one meal (one discount per table).

Question: If every couple dining there on Valentine's Day brings a coupon, what will be the mean and standard deviation of the total discounts they'll receive?

Let D = total discount (Lucky Lovers plus the coupon); then $D = X + 5$.

$$E(D) = E(X + 5) = E(X) + 5 = 5.83 + 5 = \$10.83$$
$$Var(D) = Var(X + 5) = Var(X) = 8.62^2$$
$$SD(D) = \sqrt{Var(X)} = \$8.62$$

Couples with the coupon can expect total discounts averaging $10.83. The standard deviation is still $8.62.

Multiplying each value of a random variable by a constant multiplies the mean by that constant and the variance by the square of the constant.

$E(aX) = aE(X) \qquad Var(aX) = a^2Var(X)$

$\sigma = \sqrt{a^2 Var(X)}$

$\sigma = a \cdot \text{old SD}$

On Valentine's Day at the *Quiet Nook*, couples may get a Lucky Lovers discount averaging \$5.83 with a standard deviation of \$8.62. When two couples dine together on a single check, the restaurant doubles the discount offer—\$40 for the ace of hearts on the first card and \$20 on the second.

Question: What are the mean and standard deviation of discounts for two couples dining together?

$E(2X) = 2E(X) = 2(5.83) = \$11.66$

$Var(2x) = 2^2Var(x) = 2^2 \cdot 8.62^2 = 297.2176$

$SD(2X) = \sqrt{297.2176} = \$17.24$

The expected value of the sum is the sum of the expected values

$$E(X + Y) = E(X) + E(Y)$$

The variance of the sum of two independent random variables is the sum of their individual variances.

$$Var(X \pm Y) = Var(X) + Var(Y)$$

TheValentine's Day Lucky Lovers discount for couples averages \$5.83 with a standard deviation of \$8.62. We've seen that if the restaurant doubles the discount offer for two couples dining together on a single check, they can expect to save \$11.66 with a standard deviation of \$17.24. Some couples decide instead to get separate checks and pool their two discounts.

Question: You and your amour go to this restaurant with another couple and agree to share any benefit from this promotion. Does it matter whether you pay separately or together?

Let X_1 and X_2 represent the two separate discounts, and T the total; then $T = X_1 + X_2$.

$$E(T) = E(X_1 + X_2) = E(X_1) + E(X_2) = 5.83 + 5.83 = \$11.66,$$

so the expected saving is the same either way.

The cards are reshuffled for each couple's turn, so the discounts couples receive are independent. It's okay to add the variances:

$$Var(T) = Var(X_1 + X_2) = Var(X_1) + Var(X_2) = 8.62^2 + 8.62^2 = 148.6088$$
$$SD(T) = \sqrt{148.6088} = \$12.19$$

When two couples get separate checks, there's less variation in their total discount. The standard deviation is \$12.19, compared to \$17.24 for couples who play for the double discount on a single check. It does, therefore, matter whether they pay separately or together.

The Lucky Lovers discount at the *Quiet Nook* averages \$5.83 with a standard deviation of \$8.62. Just up the street, the *Wise Fool* restaurant has a competing Lottery of Love promotion. There a couple can select a specially prepared chocolate from a large bowl and unwrap it to learn the size of their discount. The restaurant's manager says the discounts vary with an average of \$10.00 and a standard deviation of \$15.00.

How much more can you expect to save at the *Wise Fool*? With what standard deviation?

Let W = discount at the Wise Fool, X = the discount at the Quiet Nook, and D = the difference: $D = W - X$. These are different promotions at separate restaurants, so the outcomes are independent.

$$E(W - X) = E(W) - E(X) = 10.00 - 5.83 = \$4.17$$
$$\begin{aligned} SD(W - X) &= \sqrt{Var(W - X)} \\ &= \sqrt{Var(W) + Var(X)} \\ &= \sqrt{15^2 + 8.62^2} \\ &\approx \$17.30 \end{aligned}$$

Discounts at the Wise Fool will average \$4.17 more than at the Quiet Nook, with a standard deviation of \$17.30.

For random variables, does $X + X + X = 3X$? Maybe, but be careful. As we've just seen, insuring one person for $30,000 is not the same risk as insuring three people for $10,000 each. When each instance represents a different outcome for the same random variable, it's easy to fall into the trap of writing all of them with the same symbol. Don't make this common mistake. Make sure you write each instance as a *different* random variable. Just because each random variable describes a similar situation doesn't mean that each random outcome will be the same.

These are *random* variables, not the variables you saw in Algebra. Being random, they take on different values each time they're evaluated. So what you really mean is $X_1 + X_2 + X_3$. Written this way, it's clear that the sum shouldn't necessarily equal 3 times *anything*.

Egg example # 29

The *QuietNook*'s Lucky Lovers promotion offers couples discounts averaging $5.83 with a standard deviation of $8.62. The restaurant owner is planning to serve 40 couples on Valentine's Day.

$T = X_1 + X_2 + \dots + X_{40}$

What's the expected total of the discounts the owner will give? With what standard deviation?

Let $X_1, X_2, X_3, \dots, X_{40}$ represent the discounts to the 40 couples, and T the total of all the discounts. Then:

$$\begin{aligned} T &= X_1 + X_2 + X_3 + \cdots + X_{40} \\ E(T) &= E(X_1 + X_2 + X_3 + \cdots + X_{40}) \\ &= E(X_1) + E(X_2) + E(X_3) + \cdots + E(X_{40}) \\ &= 5.83 + 5.83 + 5.83 + \cdots + 5.83 \\ &= \$233.20 \end{aligned}$$

Reshuffling cards between couples makes the discounts independent, so:

$$\begin{aligned} SD(T) &= \sqrt{Var(X_1 + X_2 + X_3 + \cdots + X_{40})} \\ &= \sqrt{Var(X_1) + Var(X_2) + Var(X_3) + \cdots + Var(X_{40})} \\ &= \sqrt{8.62^2 + 8.62^2 + 8.62^2 + \cdots + 8.62^2} \\ &\approx \$54.52 \end{aligned}$$

The restaurant owner can expect the 40 couples to win discounts totaling $233.20, with a standard deviation of $54.52.

Mean and Variance for Continuous Random Variables

The mean value u_x locates the center of the continuous distribution.

The standard deviation, ~~sigma~~ σ_x, measures the extent to which the continuous distribution spreads out around u_x.

A company receives concrete of a certain type from two different suppliers.

Let x = compression strength of a randomly selected batch from Supplier 1

y = compression strength of a randomly selected batch from Supplier 2

Suppose that

μ_x = 4650 pounds/inch^2 σ_x = 200 pounds/inch^2

μ_y = 4500 pounds/inch^2 σ_y = 275 pounds/inch^2