Chapter 16
Random Variables
A random variable assumes a value based on the outcome of a random event.

- We use a capital letter, like $X$, to denote a random variable.
- A particular value of a random variable will be denoted with a lower case letter, in this case $x$. 
There are two types of random variables:
- **Discrete** random variables can take one of a finite number of distinct outcomes.
  - Example: Number of credit hours
- **Continuous** random variables can take any numeric value within a range of values.
  - Example: Cost of books this term
A probability model for a random variable consists of:

- The collection of all possible values of a random variable, and
- The probabilities that the values occur.

Of particular interest is the value we expect a random variable to take on, notated $\mu$ (for population mean) or $E(X)$ for expected value.
The expected value of a (discrete) random variable can be found by summing the products of each possible value and the probability that it occurs:

\[ \mu = E(X) = \sum x \cdot P(x) \]

Note: Be sure that every possible outcome is included in the sum and verify that you have a valid probability model to start with.
Example 1

- Find the expected value of the random variable

<table>
<thead>
<tr>
<th>x</th>
<th>P(X=x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>0.35</td>
</tr>
<tr>
<td>20</td>
<td>0.53</td>
</tr>
<tr>
<td>35</td>
<td>0.12</td>
</tr>
</tbody>
</table>

\[ E(x) = \sum x \cdot P(X) = 12(0.35) + 20(0.53) + 35(0.12) = 19 \]
A wheel comes up green 75% of the time and red 25% of the time. If it comes up green, you win $100. If it comes up red, you win nothing. Calculate the expected value of the game.

<table>
<thead>
<tr>
<th>$x$</th>
<th>$P(X=x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.75</td>
</tr>
<tr>
<td>0</td>
<td>0.25</td>
</tr>
</tbody>
</table>

\[
E(x) = \sum x \cdot P(X) = 100(0.75) + 0(0.25) = $75
\]
A company bids on two contracts. It anticipates a profit of $45,000 if it gets the larger contract and a profit of $20,000 if it gets the smaller contract. The company estimates that there is a 28% chance it will get the larger contract and a 61% chance it will get the smaller contract. If the company does not get either contract, it will neither gain nor lose money. Assuming the contracts will be awarded independently, what is the expected profit?
Example 3

\[ E(x) = \sum x \cdot P(X) = 45,000 \cdot .28 + 20,000 \cdot .61 + 0 \cdot .11 = $24,800 \]

<table>
<thead>
<tr>
<th>x</th>
<th>P(x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$45,000</td>
<td>.28</td>
</tr>
<tr>
<td>$20,000</td>
<td>.61</td>
</tr>
<tr>
<td>$0</td>
<td>.11</td>
</tr>
</tbody>
</table>
For data, we calculated the standard deviation by first computing the deviation from the mean and squaring it. We do that with discrete random variables as well.

The variance for a random variable is:

$$\sigma^2 = \text{Var}(X) = \sum (x - \mu)^2 \cdot P(x)$$

The standard deviation for a random variable is:

$$\sigma = \text{SD}(X) = \sqrt{\text{Var}(X)}$$
Example 4

Find the standard deviation of the random variable $X$.

$$Var(X) = \sum (x - \mu)^2 P(x)$$

$$\sigma = \sqrt{Var(X)}$$

$$\mu = E(X) = \sum x \cdot P(x)$$

<table>
<thead>
<tr>
<th>$x$</th>
<th>$P(X=x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>49</td>
<td>0.4</td>
</tr>
<tr>
<td>44</td>
<td>0.3</td>
</tr>
<tr>
<td>20</td>
<td>0.2</td>
</tr>
<tr>
<td>12</td>
<td>0.1</td>
</tr>
</tbody>
</table>
Example 4

\[ \mu = E(X) = \sum x \cdot P(x) = \]
\[ 49(0.4) + 44(0.3) + 20(0.2) + 12(0.1) = 38 \]

\[ \text{Var}(x) = (49 - 38)^2(0.4) + (44 - 38)^2(0.3) + (20 - 38)^2(0.2) + (12 - 38)^2(0.1) = 191.6 \]

\[ \sigma = \sqrt{\text{Var}(x)} = 13.84 \]
Example 5

In a group of 10 batteries, 3 are dead. You choose 2 batteries at random.

a) Create a probability model for the number of good batteries you get.

<table>
<thead>
<tr>
<th>Number of good</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(number of good)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Example 5

a) Create a probability model for the number of good batteries you get.

<table>
<thead>
<tr>
<th>Number of good</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(number of good)</td>
<td>.067</td>
<td>.467</td>
<td>.467</td>
</tr>
</tbody>
</table>

\[ P(0) = \frac{3}{10} \times \frac{2}{9} = \frac{6}{90} = 0.067 \]

\[ P(1) = \frac{7}{10} \times \frac{3}{9} + \frac{3}{10} \times \frac{7}{9} = 0.467 \]

\[ P(2) = \frac{7}{10} \times \frac{6}{9} = .467 \]
b) Find the expected value of the good ones you get.

\[ E(x) = \sum x P(x) = 0 \times 0.067 + 1 \times 0.467 + 2 \times 0.467 = 1.4 \]

<table>
<thead>
<tr>
<th>Number of good</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(number of good)</td>
<td>0.067</td>
<td>0.467</td>
<td>0.467</td>
</tr>
</tbody>
</table>

\[ \Sigma = \sqrt{Var(x)} = 0.61 \]

c) Find the standard deviation

\[ Var(x) = \sum (x - \mu)^2 P(x) \]
\[ = (0-1.4)^2(0.067) + (1-1.4)^2(0.467) + (2-1.4)^2(0.467) \]
\[ = 0.37416 \]
More About Means and Variances

- Adding or subtracting a constant from data shifts the mean but doesn’t change the variance or standard deviation:

\[ E(X \pm c) = E(X) \pm c \quad \text{Var}(X \pm c) = \text{Var}(X) \]

- Example: Consider everyone in a company receiving a $5000 increase in salary.
In general, 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:

\[ E(aX) = aE(X) \quad \text{Var}(aX) = a^2 \text{Var}(X) \]

Example: Consider everyone in a company receiving a 10% increase in salary.
More About Means and Variances (cont.)

- In general,
  - The mean of the sum of two random variables is the sum of the means.
  - The mean of the difference of two random variables is the difference of the means.
    \[ E(X \pm Y) = E(X) \pm E(Y) \]
  - If the random variables are independent, the variance of their sum or difference is always the sum of the variances.
    \[ Var(X \pm Y) = Var(X) + Var(Y) \]
Combining Random Variables (The Bad News)

- It would be nice if we could go directly from models of each random variable to a model for their sum.

- But, the probability model for the sum of two random variables is not necessarily the same as the model we started with even when the variables are independent.

- Thus, even though expected values may add, the probability model itself is different.
Example 6

A grocery supplier believes that in a dozen eggs, the mean number of broken eggs is 0.6 with a standard deviation of 0.5 eggs. You buy 3 dozen eggs without checking them.

a) How many broken eggs do you expect to get.
\[ E(3x) = 3E(x) = 3 \times (0.6) = 1.8 \]

b) What is the standard deviation
\[ \text{Var}(x + x + x) = .5^2 + .5^2 + .5^2 = 0.75 \]
\[ E(x + x + x) = \sqrt{0.75} = .87 \]
Continuous Random Variables

- Random variables that can take on any value in a range of values are called **continuous random variables**.
- Continuous random variables have means (expected values) and variances.
Nearly everything we’ve said about how discrete random variables behave is true of continuous random variables, as well.

When two independent continuous random variables have Normal models, so does their sum or difference.

This fact will let us apply our knowledge of Normal probabilities to questions about the sum or difference of independent random variables.
Example 6

At a certain coffee shop, all the customers buy a cup of coffee and some also buy a doughnut. The shop owner believes that the number of cups he sells each day is normally distributed with a mean of 300 cups and a standard deviation of 25 cups. He also believes that the number of doughnuts he sells each day is independent of the coffee sales and is normally distributed with a mean of 150 doughnuts and a standard deviation of 12.
a) The shop is open everyday but Sunday. Assuming day-to-day sales are independent, what is the probability he’ll sell over 2000 cups of coffee in a week?

\[
\mu = 6 \times 300 = 1800 \\
\text{Var}(6x) = 6 \times \text{Var}(x) = 6 \times 25^2 = 3750 \\
\sigma = \sqrt{3750} = 61.24 \\
Z = (2000 - 1800)/41.24 = 4.85
\]

\[
1 - N(2000, 1800, 61.24) = 0.000546 = 0.001
\]
b) If he makes a profit of 50 cents on each cup of coffee and 40 cents on each doughnut, can he reasonably expect to have a day’s profit of over $300?

\[ \text{P(profit > } $300) \]

\[ \text{Daily profit} = 0.5 \times 300 + 0.4 \times 150 = 210 \]
\[ \text{Var (0.5C) + Var(0.4D) = } 0.25 \times 25^2 + 0.16 \times 12^2 = 179.29 \]
\[ \sigma = \sqrt{179.29} = 13.39 \]

\[ Z = \frac{300 - 210}{13.39} = 6.72 \]

No. $300 is more than 6 SD away from the mean.
Example 6

c) What is the probability that on any given day he’ll sell a doughnut to more than half of his coffee customers?

Define a random variable \( Y = D - \frac{1}{2} C \). Find the probability that the random variable is greater than zero.

\[
\mu = 150 - \frac{1}{2} (300) = 0
\]

\[
\text{Var}(Y) = \text{Var}(D) + \text{Var}(0.5C) = 12^2 + 0.25 \times 25^2 = 300.25
\]

\[
\sigma = \sqrt{300.25} = 17.33
\]

\[
Z = \frac{(0 - 0)}{17.33} = 0
\]

\[
P(Z > 0) = 0.5
\]
If $X$ is a random variable with expected value $E(X)=\mu$ and $Y$ is a random variable with expected value $E(Y)=\nu$, then the covariance of $X$ and $Y$ is defined as

$$\text{Cov}(X,Y) = E((X - \mu)(Y - \nu))$$

The covariance measures how $X$ and $Y$ vary together.
Covariance, unlike correlation, doesn’t have to be between −1 and 1. If \( X \) and \( Y \) have large values, the covariance will be large as well.

To fix the “problem” we can divide the covariance by each of the standard deviations to get the correlation:

\[
\text{Corr}(X,Y) = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}
\]
What Can Go Wrong?

- Probability models are still just models.
  - Models can be useful, but they are not reality.
  - Question probabilities as you would data, and think about the assumptions behind your models.
- If the model is wrong, so is everything else.
What Can Go Wrong? (cont.)

- Don’t assume everything’s Normal.
  - You must *think* about whether the Normality Assumption is justified.

- Watch out for variables that aren’t independent:
  - You can add expected values of *any* two random variables, but
  - you can only add variances of *independent* random variables.