Plan for the week

● **M:** Indigenous People’s Day

● **W:** Probability Theory, cont’d
  ○ Random Variables
  ○ Mean & variance, revisited
  ○ Binomial, normal distributions
  ○ Law of large numbers
  ○ Central limit theorem

● **F:** No section this week!
  ○ Classical Statistics
Probability vs. Statistics

Probability
- Mathematical theory of uncertainty
- Parameters are usually known

Statistics
- The science of data analysis; techniques for making sense of data
- Parameters are usually unknown, and estimated from data
Random Variables
A baseball game is a random experiment

Sample statistics gathered:

- Number of runs, hits, and errors, per team
- Number of hits, walks, and outs, per player
- Number of strikeouts, walks, and pitches, per pitcher
- Etc.

These statistics are all examples of random variables. What are other examples of random variables?
Intuition

● A random variable is a numerical outcome of a random experiment.
● If we flip a coin $n$ times, there can be multiple random variables:
  ○ $H$ can be the random variable indicating the number of heads
  ○ $T = n - H$ can be the random variable indicating the number of tails
  ○ $X$ can represent the longest sequence of consecutive heads
  ○ $Y$ can represent the longest sequence of consecutive tails
● All these random variables are valued between 0 and $n$. 
Definition

- A **function** is a mapping from a **domain** to a **range**.
  - E.g., \( f(x,y) = x+y \) maps \( \mathbb{R}^2 \) to \( \mathbb{R} \).

- A random variable is a function:
  - The **domain** of a random variable is the sample (or outcome) space of the experiment.
    - \( \Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\} \), if we flip a coin 3 times.
  - The **range** of a random variable is a subset of \( \mathbb{R} \), the reals.

- Example: Imagine flipping a fair coin 2 times.
  - Let \( H \) be the random variable representing the total number of heads.
  - Q: What is the domain? I.e. what are the possible outcomes? A: \( \{HH, HT, TH, TT\} \).
  - Q: What is the range of \( H \)? I.e., what are the possible values of \( H \)? A: \( \{0, 1, 2\} \).
  - Thus, \( H \) is a function from \( \{HH, HT, TH, TT\} \) to \( \{0, 1, 2\} \).
Inversely, a random variable describes an event: i.e., a set of outcomes.

The probability distribution associated with a random variable is determined by the probabilities of the corresponding event.

For example:

- $\Pr[H = 0] = \frac{|\{TT\}|}{|\{HH, HT, TH, TT\}|} = \frac{1}{4}$
- $\Pr[H = 1] = \frac{|\{HT, TH\}|}{|\{HH, HT, TH, TT\}|} = \frac{1}{2}$
- $\Pr[H = 2] = \frac{|\{HH\}|}{|\{HH, HT, TH, TT\}|} = \frac{1}{4}$

All probabilities are between 0 and 1, and together they sum to 1.
Random Variables Vary

- What makes random variables interesting is the fact that they vary!
- You can and will get different results running the same random experiment over and over again, as a result of the randomness.
- This randomness is described by a probability distribution, which is often summarized by its center and spread.
Expectation
Expected Value

- Once we have a specified the range of a random variable, and its probability distribution, we can calculate the expected value of that random variable.
- Assume $X$ is a random variable, $x_i$ is an element of its range, and $p_i$ is the probability of $x_i$. Then the expected value of $X$ is calculated as follows:

$$E(X) = \sum_{i=1}^{n} x_ip_i = x_1p_1 + x_2p_2 + x_3p_3 + \cdots + x_n p_n$$

- $E[H] = (0)Pr[H = 0] + (1)Pr[H = 1] + (2)Pr[H = 2] = 0(\frac{1}{4}) + 1(\frac{1}{2}) + (2)(\frac{1}{4}) = 1$
  - If we flip a coin two times, we should expect to see one head.
- The expected value of a random variable is also called the mean ($\mu$).
If $X$ is a random variable representing the result of one roll of a six-sided die, then what is the expected value of $X$?

A) 3  
B) 3.5  
C) 4
iClicker Answer

If $X$ is a random variable for the number rolled on a six-sided die, then what is the expected value of $X$?

A) 3
B) 3.5
C) 4

$\frac{1}{6} (1 + 2 + 3 + 4 + 5 + 6) = 3.5$

This is the expected value, even though we cannot roll a 3.5 ever!
Law of Large Numbers

- You can roll a die many times.
- Each time, you will see a 1, 2, 3, 4, 5, or 6.
- If you average the value you see across all your trials, as you run more and more experiments, this average will approach 3.5!

<table>
<thead>
<tr>
<th>Trial</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>6</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Average</td>
<td>1</td>
<td>2.5</td>
<td>8/3</td>
<td>11/4</td>
<td>16/5</td>
<td>22/6</td>
<td>24/7</td>
<td>27/8</td>
<td>31/9</td>
<td>32/10</td>
<td>38/11</td>
<td>44/12</td>
</tr>
</tbody>
</table>

- In this experiment, the **sample mean** is 3.667.
Law of Large Numbers

- You can toss 2 coins many times.
- Each time, you will see 0, 1, or 2 heads.
- If you average the number of heads you see across all your trials, as you run more and more experiments, this average will approach 1!

<table>
<thead>
<tr>
<th>Trial</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Average</td>
<td>0</td>
<td>1</td>
<td>4/3</td>
<td>1</td>
<td>1</td>
<td>5/6</td>
<td>5/7</td>
<td>6/8</td>
<td>10/9</td>
<td>12/10</td>
<td>12/11</td>
<td>13/12</td>
</tr>
</tbody>
</table>

- In this experiment, the sample mean is 1.083.
Variance
Recall Sample Variance

Assume $X$ is a random variable, $x_i$ is an element of its range, and $\mu$ is its mean. The sample variance of $X$, given a sample of size $N$, is calculated as follows:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2$$

<table>
<thead>
<tr>
<th>Trial</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>6</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Squared Differences</td>
<td>6.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>2.25</td>
<td>6.25</td>
<td>2.25</td>
<td>0.25</td>
<td>0.25</td>
<td>6.25</td>
<td>6.25</td>
<td>0.25</td>
</tr>
</tbody>
</table>

- In this experiment, the sample variance is 2.58, and the sample SD is 1.61.
Recall Sample Variance

Assume $X$ is a random variable, $x_i$ is an element of its range, and $\mu$ is its mean. The sample variance of $X$, given a sample of size $N$, is calculated as follows:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2$$

<table>
<thead>
<tr>
<th>Trial</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Squared Differences</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

- In this experiment, the sample variance is .75, and the sample SD is .86.
Variance through the Prism of Expectation

- Actually, variance is not defined as the average squared deviation, but rather as the expected squared deviation.
- In the case of fair die — actually, any time probabilities are uniform — average and expectation are the same thing:
  - \(\frac{1}{6}(1 - 3.5)^2 + \frac{1}{6}(2 - 3.5)^2 + \frac{1}{6}(3 - 3.5)^2 + \frac{1}{6}(4 - 3.5)^2 + \frac{1}{6}(5 - 3.5)^2 + \frac{1}{6}(6 - 3.5)^2 \approx 2.9\)
  - The sum of the squared deviations is \((1 - 3.5)^2 + (2 - 3.5)^2 + (3 - 3.5)^2 + (4 - 3.5)^2 + (5 - 3.5)^2 + (6 - 3.5)^2 = 17.5\). To find the average squared deviation, we divide by 17.5/6 \(\approx 2.9\)
- But probabilities are not always uniform. If our random variable is the sum of two dice, 7 is a more likely outcome than 12.
- Variance, more generally, is defined as an expectation.
Variance through the Prism of Expectation

- Once again, variance is the expected value of the squared deviations between the values of a random variable and its mean.
  - Variance$[X] = E[(X - E[X])^2]$
  - We can simplify this through some algebra.
  - Variance$[X] = E[X^2] - E[X]^2$

- This second formula will yield the same result as the first, and it is usually easier to work with! The variance of the sum of two dice is:
  - $E[X^2] = (\frac{1}{36})2^2 + (\frac{1}{18})3^2 + (\frac{1}{12})4^2 + (\frac{1}{9})5^2 + (\frac{5}{36})6^2 + (\frac{5}{36})7^2 + (\frac{5}{36})8^2 + (\frac{9}{9})9^2 + (\frac{1}{12})10^2 + (\frac{1}{18})11^2 + (\frac{1}{36})12^2 = 54.83$
  - $E[X]^2 = (7)^2 = 49$
  - $E[X^2] - E[X]^2 = 54.83 - 49 \sim 5.83$
Recall Sample Covariance

Assume $X$ is a random variable, $x_i$ is an element of its range, and $\mu_x$ is its mean. Assume $Y$ is a random variable, $y_i$ is an element of its range, and $\mu_y$ is its mean. The sample covariance of $X$ and $Y$, given a sample of size $N$, is calculated as follows:

$$\gamma_{XY} = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu_x)(y_i - \mu_y)$$

$$c_{XY} = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})$$

Actually, covariance is not defined as the average product of the deviations, but rather as the expected product.
Covariance through the Prism of Expectation

- Once again, covariance is the product of the expected values of the deviations between the values of two random variables and their means.
  - Covariance\[X, Y\] = \(E[(X - E[X])(Y - E[Y])]\)
  - We can simplify this through some algebra.
  - Covariance\[X, Y\] = \(E[X, Y] - E[X]E[Y]\)

- This second formula will yield the same result as the first, and it is usually easier to work with! The covariance of rain today and rain tomorrow is:
  - \(E[X, Y] = .55(X = 1, Y = 1) + .05(X = 1, Y = 0) + .2(X = 0, Y = 1) + .2(X = 0, Y = 0) = .55\)
  - \(E[X] = P(X = 1) = .6\)
  - \(E[Y] = P(Y = 1) = .75\)
  - \(E[X, Y] - E[X]E[Y] = .55 - .45 = .1\)

- Corr\[X, Y\] = Cov\[X, Y\] / \(\sigma_X \sigma_Y\)

### Joint Distribution

<table>
<thead>
<tr>
<th>JOINT DISTRIBUTION</th>
<th>Rain tomorrow (.75)</th>
<th>No rain tomorrow (.25)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rain today (.6)</td>
<td>.55</td>
<td>.05</td>
</tr>
<tr>
<td>No rain today (.4)</td>
<td>.2</td>
<td>.2</td>
</tr>
</tbody>
</table>
A Technical Aside
The Linearity of Expectation

- We introduced the concept of expectation in the last lecture.
- Let $X_1$ be a random variable represented the first die roll.
  - $E[X_1] = 3.5$
- Now, let $X_2$, $X_3$, $X_4$, and $X_5$ represent additional die rolls.
  - $E[X_1+X_2+X_3+X_4+X_5] = E[5X_1] = 5E[X_1] = 17.5$
- More generally:
  - $E[aY_1 + bY_2] = \color{blue}{E[aY_1] + E[bY_2]} = aE[Y_1] + bE[Y_2]$
  - This rule is called the linearity of expectation.
The Linearity of Standard Deviation

- The same property of linearity holds for standard deviation as well.
  - \( \text{SD}[aY_1 + bY_2] = a \text{SD}[Y_1] + b \text{SD}[Y_2] \)

- Remember that variance is the square of standard deviation.
  - \( \text{Var}[aY_1 + bY_2] = a^2 \text{Var}[Y_1] + b^2 \text{Var}[Y_2] \)

- Summary:
  - You can pull constants out of expectation and standard deviation formulas.
  - You can pull constants out of variance formulas, but then you must square the constant when you pull it out.

- One more important note:
  - \( \text{Var}[X - Y] = \text{Var}[X] + \text{Var}[Y] \)
  - Why? \( E[(X-Y)^2] - (E[X-Y])^2 \Rightarrow \text{simplify and you’ll get this result!} \)