Class 11

DATA1220-55, Fall 2024

Sarah E. Grabinski

2024-09-23

The General Addition Rule

The probability of event A ***or*** event B occurring is the sum of the probability that A occurs and the probability that B occurs minus the probability that A *and* B occurs.

$$\begin{aligned} P(A \,\text{or}\, B) &= P(A) + P(B) - P(A \,\text{and}\, B) \\ &= P(A) + P(B) - P(A \cup B) \\ &= P(A \cap B) \end{aligned}$$

Example: The General Addition Rule for Non-Disjoint Events

- The LA Dodgers have played 156 games so far in the 2024 season.
- Shohei Hotani has hit 53 home runs and stolen 55 bases so far in the 2024 season. He did both in one game 34 times.
- What's the probability that Shohei Hotani hits a home run *or* steals a base during the next Dodgers game?

Describing the sample space

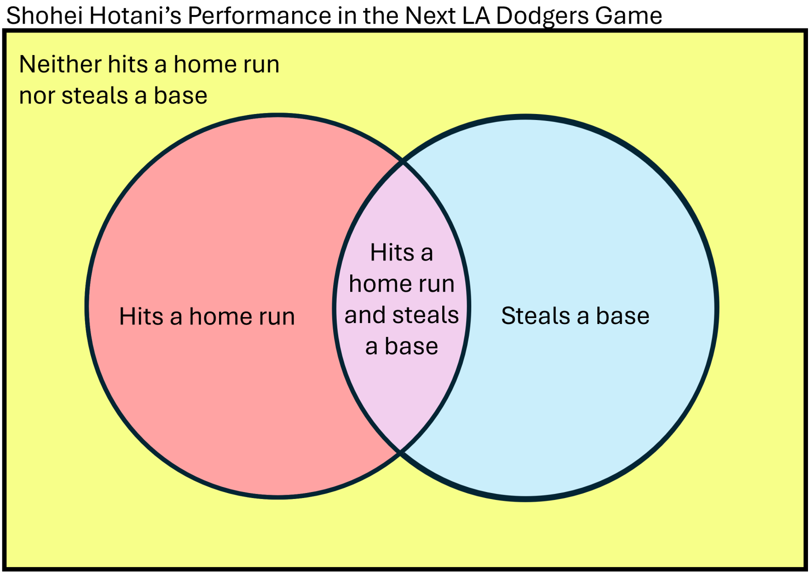

Figure 1: The sample space for Shohei Otani's performance during the

Zooming in

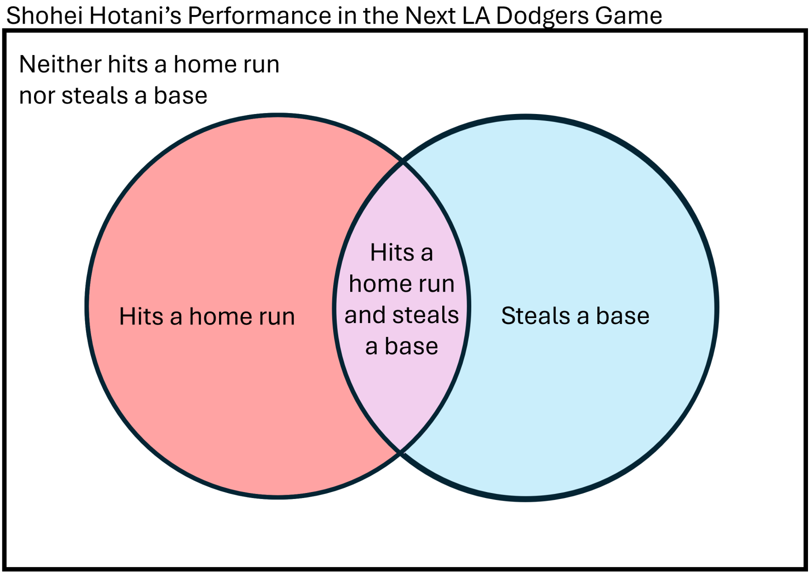

Figure 2: The probability that Shohei Otani hits a home run ***or*** steals a

What's the probability that Shohei Otani hits a home run during the next game?

Shohei Hotani hits a home run approximately 1 out of every 3 games.

$$\begin{aligned} P(\text{hitsahomerun}) &= \frac{\text{count}(\text{homeruns})}{\text{count}(\text{totalgames})} \\ &= \frac{53}{156} \\ &= 0.340 \end{aligned}$$

What's the probability that Shohei Otani steals a base during the next game?

Shohei Hotani steals a base approximately 1 out of every 3 games.

$$
\begin{aligned}
P(\text{steals a base}) &= \frac{\text{count}(\text{stolen bases})}{\text{count}(\text{total games})} \\
&= \frac{55}{156} \\
&= 0.353
\end{aligned}
$$

What's the probability that Shohei Otani hits a home run *and* steals a base during the next game?

Shohei Hotani hits a home run *and* steals a base approximately 1 out of every 4-5 games.

$$
\begin{aligned}
P\begin{pmatrix}\text{homerunand}\\ \text{stolenbase}\end{pmatrix} &= \frac{\text{count}\begin{pmatrix}\text{homerunand}\\ \text{stolenbase}\end{pmatrix}}{\text{count}(\text{totalgames})} \\
&= \frac{34}{156} \\
&= 0.218
\end{aligned}
$$

Putting it all together

Shohei Hotani hits a home run or steals a base approximately 1 out of every 2 games.

$$
\begin{aligned}
P\begin{pmatrix}\text{homerunor}\\ \text{stolenbase}\end{pmatrix} &= P(\text{homerun}) + P(\text{stolenbase}) \\
&\qquad - P\begin{pmatrix}\text{homerunand}\\ \text{stolenbase}\end{pmatrix} \\
&= \frac{53}{156} + \frac{55}{156} - \frac{34}{156} \\
&= 0.474
\end{aligned}
$$

The Odds

- The ***odds*** of an event occurring is the probability of an event occurring divided by the probability of the event not occurring (i.e. its complement).
- The odds are *for* or *in favor* of event A if $P(A) > P(A')$ $(P(A) > 0.5)$ and $\text{odds} > 1$
- The odds *disfavor* or are *against* event A if $P(A) < P(A')$ $(P(A) < 0.5)$ and $\text{odds} < 1$.

Describing the odds of an event

- The odds of event A are expressed as a simplified ratio in terms of positive integers a and a' where a is the approximate number of times that event A occurs for every time a' that event A does not occur.
  - $a : a'$ in favor of event A, when \$P(A) $>$ P(A') and $a > a'$
  - $a' : a$ against event A, when \$P(A) $<$ P(A') and $a' > a$
- Example: if $P(A) = 0.9$ and $\text{odds}(A) = 9$, then you would say, "The odds are 9:1 in favor of event A."
- Example: if $P(A) = 0.1$ and $\text{odds}(A) = 0.111$, then you would say, "The odds are 9:1 against event A."

Example: The Odds

If the probability that Shohei Otani hits a home run or steals a base in the next game is 0.474, then the odds against Shohei Otani hitting a home run or stealing a base are ~11:10 (i.e. about even).

$$\begin{aligned}
\text{odds}\begin{pmatrix}\text{homerunor}\\ \text{stolenbase}\end{pmatrix} &= \frac{1 - P\begin{pmatrix}\text{homerunor}\\ \text{stolenbase}\end{pmatrix}}{P\begin{pmatrix}\text{homerunor}\\ \text{stolenbase}\end{pmatrix}} \\
&= \frac{1 - 0.474}{0.474} \\
&= 1.11
\end{aligned}$$

The Addition Rule for Disjoint Events

When events A and B are ***disjoint***, the probability of event A ***or*** event B occurring is just the sum of the probability that A occurs and the probability that B occurs, because the probability that event A *and* event B occurs is 0.

$$\begin{aligned} P(A \,\text{or}\, B) &= P(A) + P(B) - P(A \,\text{and}\, B) \\ &= P(A) + P(B) \\ &= P(A \cap B) \end{aligned}$$

Example: The Addition Rule for Disjoint Events

- Shohei Otani has had 611 at bats far in the 2024 season.
- Shohei Hotani has hit 53 home runs and stolen 55 bases so far in the 2024 season.
- What's the probability that Shohei Hotani hits a home run *or* steals a base during his next at bat?

Describing the sample space

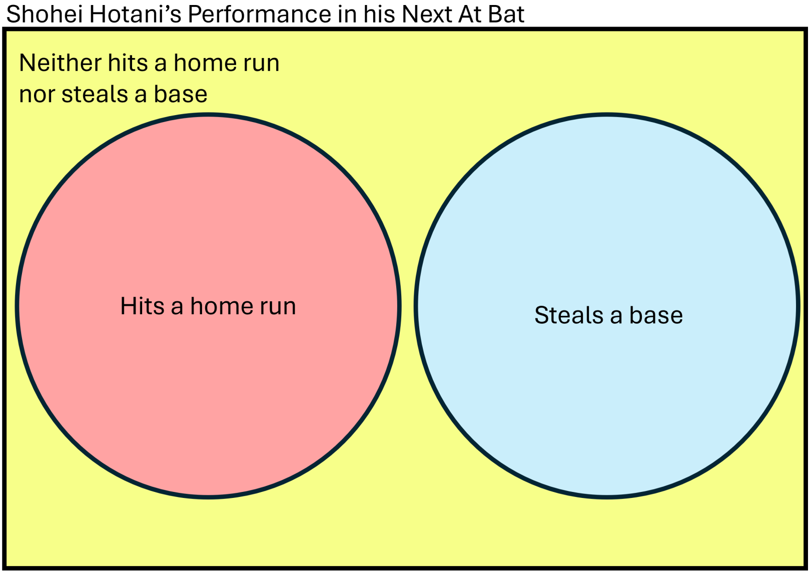

Figure 3: The sample space for Shohei Otani's performance during his

Zooming in

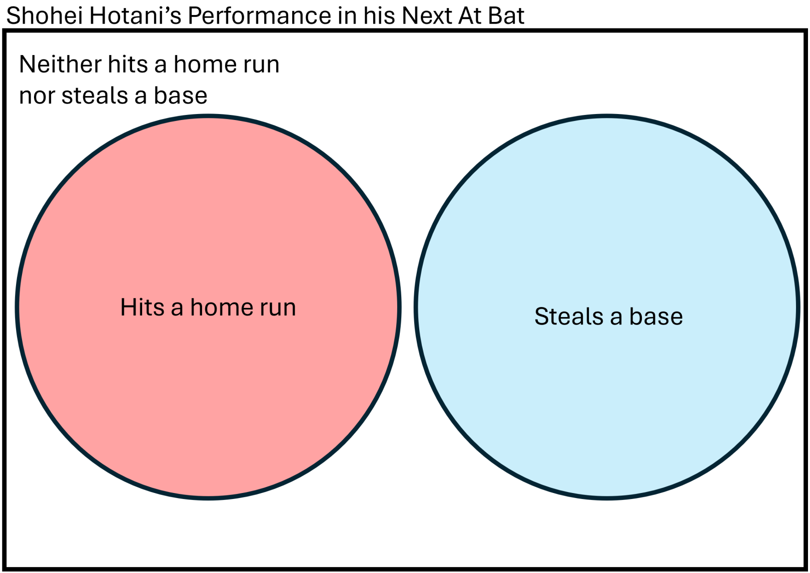

Figure 4: The probability that Shohei Otani hits a home run ***or*** steals a

What's the probability that Shohei Otani hits a home run during his next at bat?

Shohei Hotani hits a home run approximately 1 out of every 11 at bats.

$$
\begin{aligned}
P(\text{hitsahomerun}) &= \frac{\text{count}(\text{homeruns})}{\text{count}(\text{atbats})} \\
&= \frac{53}{611} \\
&= 0.087
\end{aligned}
$$

What's the probability that Shohei Otani steals a base during his next at bat?

Shohei Hotani steals a base approximately 1 out of every 11 at bats.

$$
\begin{aligned}
P(\text{stealsabase}) &= \frac{\text{count}(\text{stolenbases})}{\text{count}(\text{atbats})} \\
&= \frac{55}{611} \\
&= 0.090
\end{aligned}
$$

Putting it all together

Shohei Hotani hits a home run or steals a base approximately 1 out of every 5 at bats.

$$\begin{aligned} P\begin{pmatrix}\text{homerunor} \\ \text{stolenbase}\end{pmatrix} &= P(\text{homerun}) + P(\text{stolenbase}) \\ &= \frac{53}{611} + \frac{55}{611} \\ &= 0.177 \end{aligned}$$

Dependent Processes

- If random process B is ***dependent*** on random process A, then the probability of random process B varies based on the outcome of random process A
- *i.e. knowing the outcome of A provides additional information about the probability of B*
- Example: When listening to a playlist using a "modern shuffle", the probability that the next song will be by a particular artist *does* change based on whether or not the last song played was also by that artist.

The General Multiplication Rule

The probability of event A ***and*** event B occurring is the product of the probability that A occurs and the *conditional probability* that B occurs given that A has already occurred.

$$
\begin{aligned}
P(A \,\text{and}\, B) &= P(A) \times P(B \,\text{given}\, A) \\
&= P(A) \times P(B|A) \\
&= P(A \cup B)
\end{aligned}
$$

Independent Processes

- If random process B is ***independent*** of random process A, then the probability of random process B does NOT vary based on the outcome of random process A
- *i.e. knowing the outcome of A does NOT provide additional information about the probability of B*
- Example: When listening to a playlist using a "true shuffle", the probability that the next song will be by a particular artist *does not* change based on whether or not the last song played was also by that artist.

Multiplication Rule for Independent Processes

The probability of event A ***and*** event B occurring is the product of the probability that A occurs and the probability that B occurs, because the probability of B does not change based on the outcome of A.

$$\begin{aligned} P(A \,\mathrm{and}\, B) &= P(A) \times P(B \,\mathrm{given}\, A) \\ &= P(A) \times P(B|A) \\ &= P(A) \times P(B) \\ &= P(A \cup B) \end{aligned}$$

How do you know if two random processes are independent?

- Compare the conditional probabilities of B given the different possible outcomes of A. If $P(B|A) \approx P(B)$ for all values of A, then the two random processes are likely independent.
- Calculate the probability that event A and B occur under both an independence model ($P(A \,\text{and}\, B) = P(A) \times P(B)$) and a dependence model ($P(A \,\text{and}\, B) = P(A) \times P(B|A)$.
  - If $P(A) \times P(B) \approx P(A) \times P(B|A)$, then A and B are likely ***independent processes***.
  - If $P(A) \times P(B) \neq P(A) \times P(B|A)$, then A and B are likely ***dependent*** processes.

Example 1: Checking for Independence

- The LA Dodgers have played 156 games so far in the 2024 season.
- Shohei Hotani has hit 53 home runs and stolen 55 bases so far in the 2024 season. In the 49 games in which he has hit a home run, he has stolen 18 bases.
- Is the random process of Shohei Hotani hitting a home run ***independent*** of the random process of Shohei Hotani stealing a base?

Describing the sample space

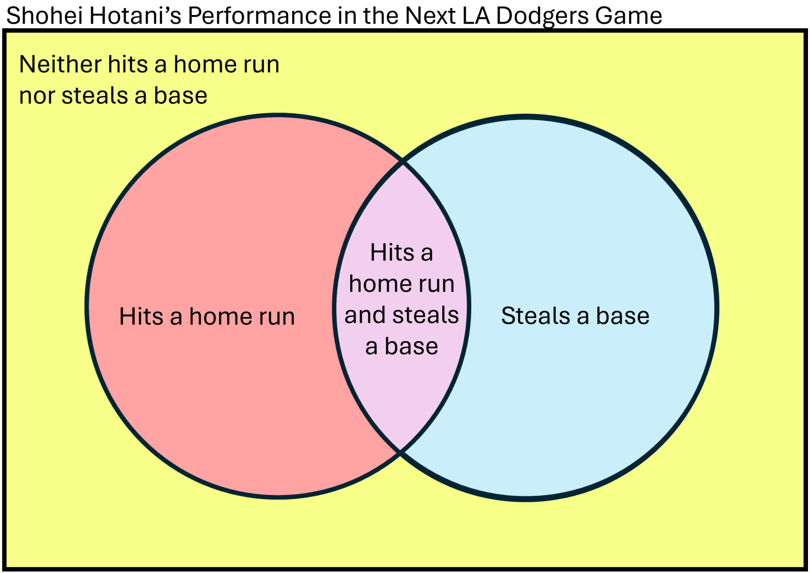

Figure 5: The sample space for Shohei Otani's performance during the

Zooming in

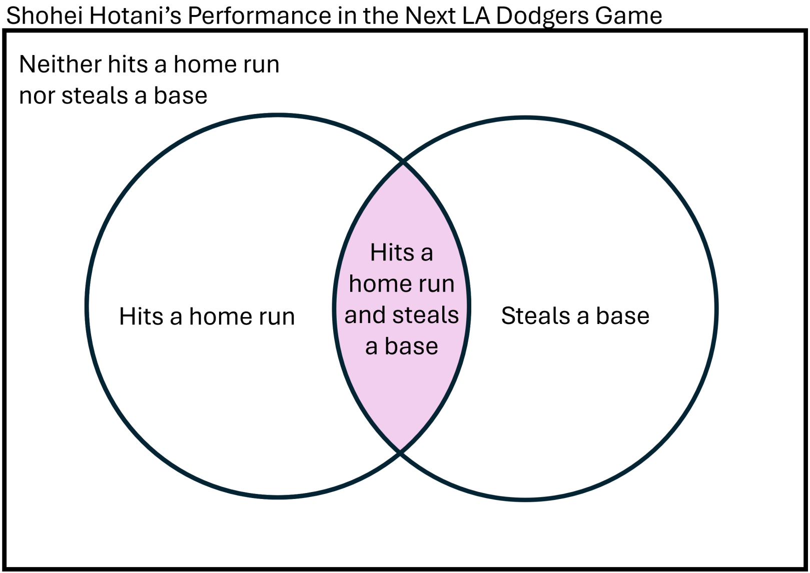

Figure 6: The probability that Shohei Otani hits a home run ***and*** steals a

What's the probability that Shohei Otani hits a home run during the next game?

Shohei Hotani hits a home run approximately 1 out of every 3 games.

$$\begin{aligned} P(\text{hitsahomerun}) &= \frac{\text{count(homeruns)}}{\text{count(totalgames)}} \\ &= \frac{53}{156} \\ &= 0.340 \end{aligned}$$

What's the conditional probability that Shohei Otani steals a base *given* that he has hit a home run during the game?

Shohei Hotani steals a base approximately 1 out of every 3-4 games in which he hits a one home run.

$$
\begin{aligned}
P\begin{pmatrix} \text{stolenbase} \\ \text{givenhomerun} \end{pmatrix} &= \frac{\text{count}(\text{stolenbases})}{\text{count}(\text{homerungames})} \\
&= \frac{18}{49} \\
&= 0.286
\end{aligned}
$$

Putting it all together

Shohei Otani hits a home run ***and*** steals a base approximately 1 out of every 8 games.

$$
\begin{aligned}
P\begin{pmatrix}\text{homerunand}\\ \text{stolenbase}\end{pmatrix} &= P(\text{homerun}) \\
&\quad \times P(\text{stolenbase} \mid \text{homerun}) \\
&= \frac{53}{156} \times \frac{18}{49} \\
&= 0.125
\end{aligned}
$$

What's the probability that Shohei Otani steals a base given he does *NOT* hit a home run during the next game?

Shohei Hotani steals a base approximately 1 out of every 3 games in which he does *NOT* hit a home run.

$$
\begin{aligned}
P\begin{pmatrix}\text{stolenbasegiven}\\ \text{nohomerun}\end{pmatrix} &= \frac{\text{count}(\text{stolenbases})}{\text{count}(\text{nohomerungames})}\\
&= \frac{55-18}{156-49}\\
&= 0.346
\end{aligned}
$$

Comparing conditional probabilities

- The conditional probability $P(\mathrm{stolenbase} \,|\, \mathrm{homerun})$ that Shohei Otani steals a base during the next Dodgers game given that he has scored a home run is 0.286.
- The conditional probability $P(\mathrm{stolenbase} \,|\, \mathrm{nohomerun})$ that Shohei Otani steals a base during the next Dodgers game given that he has not scored a home run is 0.346.

Is the probability that Shohei Otani steals a base during a game meaningfully different based on whether or not he has hit a home run?

Calculating joint probability under a model of independence

$$
\begin{aligned}
P\begin{pmatrix}\text{homerunand}\\ \text{stolenbase}\end{pmatrix} &= P(\text{homerun}) \times P(\text{stolenbase}) \\
&= \frac{53}{156} \times \frac{55}{156} \\
&= 0.353
\end{aligned}
$$

Comparing probability models

- The probability that Shohei Otani steals a base and hits a home run during the next Dodgers game under a model of independence is 0.353.
- The probability that Shohei Otani steals a base and hits a home run during the next Dodgers game under a model of dependence a home run is 0.125

Is the probability that Shohei Otani steals a base and hits a home run under a model of independence meaningfully different from the model of dependence?

Example 2: Checking for Independence

- The LA Dodgers have played 156 games so far in the 2024 season.
- Shohei Hotani has made at least one hit in 107 games and pitched at least one strikeout in 105 games so far in the 2024 season. He did both in 75 games.
- Is the random process of Shohei Hotani making at least one hit ***independent*** of the random process of Shohei Hotani pitching at least one strikeout?

Describing the sample space

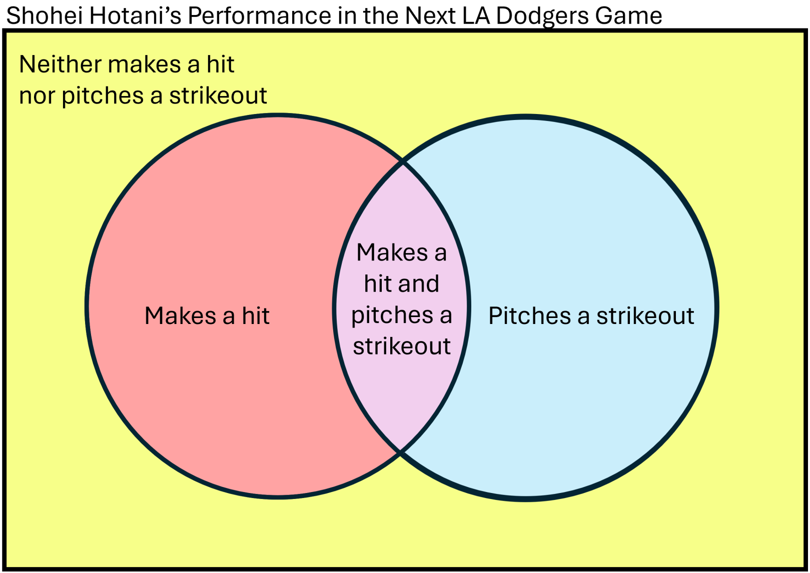

Figure 7: The sample space for Shohei Otani's performance during the

Zooming in

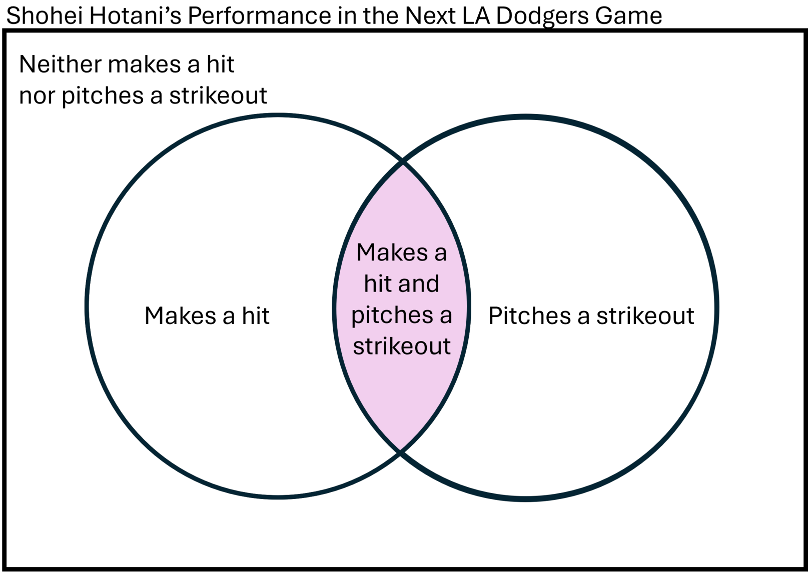

Figure 8: The probability that Shohei Otani makes at least one hit ***and***

What's the probability that Shohei Otani makes 1+ hits during the next game?

Shohei Hotani makes at least one hit approximately 2 out of every 3 games.

$$
\begin{aligned}
P(\text{hits} > 0) &= \frac{\text{count}(\text{hits} > 0\,\text{games})}{\text{count}(\text{totalgames})} \\
&= \frac{107}{156} \\
&= 0.686
\end{aligned}
$$

What's the probability that Shohei Otani pitches 1+ strikeouts *given* he makes 1+ hits?

Shohei Hotani pitches at least one strikeout approximately 2 out of every 3 games in which he makes at least one hit.

$$P\begin{pmatrix}\text{strikeouts} > 0\\ \text{givenhits} > 0\end{pmatrix} = \frac{\text{count}\begin{pmatrix}\text{strikeouts} > 0,\\ \text{hits} > 0\,\text{games}\end{pmatrix}}{\text{count}(\text{hits} > 0\,\text{games})}$$

$$= \frac{75}{107}$$

$$= 0.701$$

Putting it all together

Shohei Otani makes at least one hit ***and*** pitches at least one strikeout approximately 1 out of every 2 games.

$$
\begin{aligned}
P\begin{pmatrix}\text{hits} > 0\,\text{and} \\ \text{strikeouts} > 0\end{pmatrix} &= P(\text{hits} > 0) \\
&\quad \times P\begin{pmatrix}\text{strikeouts} > 0 \\ \text{givenhits} > 0\end{pmatrix} \\
&= \frac{107}{156} \times \frac{75}{107} \\
&= 0.481
\end{aligned}
$$

What's the probability that Shohei Otani pitches at least one strikeout if he does *NOT* make at least one hit during the next game?

Shohei Hotani pitches a strikeout approximately 2 out of every 3 games in which he does *NOT* make a hit.

$$
\begin{aligned}
P\begin{pmatrix}\text{strikeouts} > 0\\ \text{given} \text{hits} = 0\end{pmatrix} &= \frac{\text{count}\begin{pmatrix}\text{strikeouts} > 0,\\ \text{hits} = 0\,\text{games}\end{pmatrix}}{\text{count}(\text{hits} = 0\,\text{games})}\\
&= \frac{107 - 75}{156 - 107}\\
&= 0.653
\end{aligned}
$$

Comparing conditional probabilities

- The probability $P(\text{strikeouts} > 0|\,\text{hits} > 0)$ that Shohei Otani pitches a strikeout given that he has made a hit is 0.701.
- The probability $P(\text{strikeouts} > 0|\,\text{hits} = 0)$ that Shohei Otani steals a base during the next Dodgers game given that he has not scored a home run is 0.653.

Is the probability that Shohei Otani pitches a strikeout during a game meaningfully different based on whether or not he makes a hit?

Calculating joint probability under a model of independence

$$
\begin{aligned}
P\left(\begin{matrix}\text{hits} > 0,\\ \text{strikeouts} > 0\end{matrix}\right) &= P(\text{hits} > 0) \times P(\text{strikeouts} > 0)\\
&= \frac{107}{156} \times \frac{105}{156}\\
&= 0.462
\end{aligned}
$$

Comparing probability models

- The probability that Shohei Otani pitches a strikeout and makes a hit during the next game under a model of independence is 0.462.
- The probability that Shohei Otani steals a base and hits a home run during the next Dodgers game under a model of dependence a home run is 0.481

Is the probability that Shohei Otani pitches a strikeout and makes a hit during the game under a model of independence meaningfully different from the model of dependence?

How do we decide when a difference is meaningful?

That's what statistics is all about!