Introduction to Statistics

Nels Grevstad

Metropolitan State University of Denver

ngrevsta@msudenver.edu

September 2, 2019

Topics

Objectives

Objectives:

- Identify the sample space of a chance experiment.
- Interpret an event as a group of outcomes in the sample space.
- Interpret probabilities as long-run proportions, and know the properties that all probabilities satisfy.
- Compute probabilities of events when outcomes are equally likely.

(objectives cont'd):

- Form new events using ”not”, ”and”, and ”or”.
- Recognize mutually exclusive events.
- Produce a Venn diagram.
- Apply the Complementation Rule and the Special Addition Rule to obtain probabilities.
- Apply the General Addition Rule to obtain a probability.

Probability Basics, Events (4.1, 4.2)

Chance Experiments, Sample Spaces, and Events

- A ***chance experiment*** is any action or process whose **outcome is uncertain.**

Probability Basics, Events (4.1, 4.2)

Chance Experiments, Sample Spaces, and Events

- A ***chance experiment*** is any action or process whose **outcome is uncertain.**

Example

Some examples of **chance experiments**:

- Tossing a coin

Probability Basics, Events (4.1, 4.2)

Chance Experiments, Sample Spaces, and Events

- A ***chance experiment*** is any action or process whose **outcome is uncertain.**

Example

Some examples of **chance experiments**:

- Tossing a coin
- Rolling a six-sided die

Probability Basics, Events (4.1, 4.2)

Chance Experiments, Sample Spaces, and Events

- A ***chance experiment*** is any action or process whose **outcome is uncertain.**

Example

Some examples of **chance experiments**:

- Tossing a coin
- Rolling a six-sided die
- Randomly selecting a student from a university

Probability Basics, Events (4.1, 4.2)

Chance Experiments, Sample Spaces, and Events

- A ***chance experiment*** is any action or process whose **outcome is uncertain.**

Example

Some examples of **chance experiments**:

- Tossing a coin
- Rolling a six-sided die
- Randomly selecting a student from a university
- Drawing a card from a shuffled deck

- The ***sample space*** of a chance experiment is the **collection of all the possible outcomes.**

- The ***sample space*** of a chance experiment is the **collection of all the possible outcomes.**

Example

Here are the **sample spaces** for some of the chance experiments of the last example.

- If we toss a coin, the **sample space** consists of the two outcomes, heads and tails, which we'll write as

$$\textbf{Sample Space} = \{ H, T \}.$$

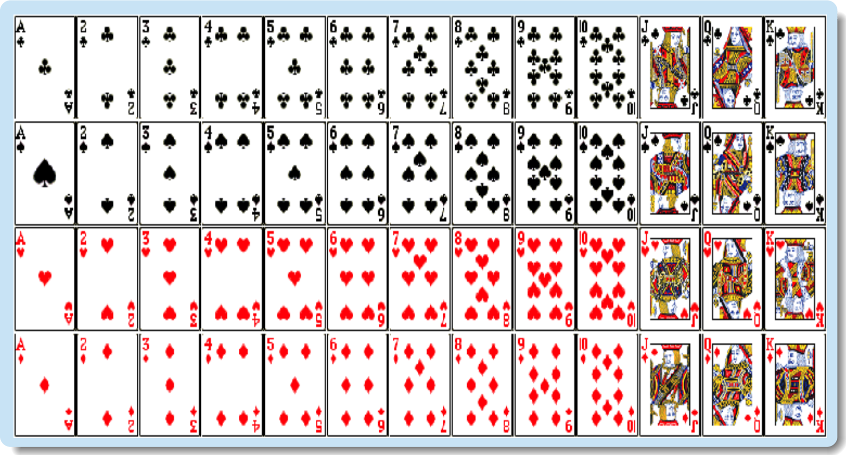
- If we roll a six-sided die, the **sample space** is comprised of the six outcomes, which we write as

$$\textbf{Sample Space} = \{⚀, ⚁, ⚂, ⚃, ⚄, ⚅\}.$$

- If we randomly select a student from a university, the **sample space** is the entire list of students at the university (i.e. the set of ”names in the hat”):

$$\textbf{Sample Space} = \{\text{Stephanie Lawson}, \\ \text{Jeffrey Miller}, \\ \text{Angela DuPont}, \\ \vdots \\ \text{Melissa Jordan} \\ \text{Karl Stephenson}\}$$

- If we draw a card from a deck, the **sample space** is the entire set of 52 cards in the deck, as shown on the next slide.



- An ***event*** is **a single outcome or a group of outcomes** from within the sample space.

- An ***event*** is **a single outcome or a group of outcomes** from within the sample space.

  We'll use **capital letters** (e.g. A, B, C, E, F, etc.) to denote **events**.

- An ***event*** is **a single outcome or a group of outcomes** from within the sample space.

  We'll use **capital letters** (e.g. A, B, C, E, F, etc.) to denote **events**.

Example

Here are some **events** associated with the chance experiments of the previous two examples.

- An ***event*** is **a single outcome or a group of outcomes from within the sample space.**

  We'll use **capital letters** (e.g. A, B, C, E, F, etc.) to denote **events.**

Example

Here are some **events** associated with the chance experiments of the previous two examples.

- Toss a coin. Consider the **event A** defined as

$$A = \text{The event that the coin lands on heads.}$$

- An ***event*** is **a single outcome or a group of outcomes** from within the sample space.

  We'll use **capital letters** (e.g. A, B, C, E, F, etc.) to denote **events**.

Example

Here are some **events** associated with the chance experiments of the previous two examples.

- Toss a coin. Consider the **event A** defined as

  A = The event that the coin lands on heads.

  Then **A** consists of the single outcome, which we write as

  $$A = \{H\}.$$

- Roll a six-sided die. Consider the **event B** defined as

$$B = \text{The event that the die lands on an even number.}$$

- Roll a six-sided die. Consider the **event B** defined as

  B = The event that the die lands on an even number.

  Then **B** consists of the group of the three outcomes

  B = {⚁, ⚃, ⚅}

- Randomly choose a student from a university, and define the **event E** to be

$$E = \text{The event that the student is female.}$$

- Randomly choose a student from a university, and define the **event E** to be

$$E = \text{The event that the student is female.}$$

Then **E** consists of the group of all the female students at the university, i.e.

$$\begin{aligned} E &= \{\text{Stephanie Lawson}, \\ &\quad \text{Angela DuPont}, \\ &\quad \vdots \\ &\quad \text{Melissa Jordan}\} \end{aligned}$$

Definition of Probability

- The ***probability*** of an event A, denoted **P(A)**, is defined to be the **long-run proportion** of times A occurs.

Definition of Probability

- The ***probability*** of an event A, denoted **P(A)**, is defined to be the **long-run proportion** of times A occurs.

  So, upon repeating the chance experiment a **very large** number of times,

$$P(A) = \frac{\text{Number of times A occurs}}{\text{Number of times the chance experiment is repeated}}$$

Example

Toss a coin repeatedly. In the **long-run**, the coin will land H (heads) **50%** the time. So

$$P\left(H\right) = \mathbf{0.5}$$

Probability for Equally Likely Outcomes

- When the outcomes in the sample space are **equally likely**, **P(A)** is just the **fraction** of outcomes that comprise A.

Probability for Equally Likely Outcomes

- When the outcomes in the sample space are **equally likely**, **P(A)** is just the **fraction** of outcomes that comprise A.

Probability for Equally Likely Outcomes (f/N Rule)

Suppose an experiment has N possible outcomes, all equally likely. An event A that's comprised of f of the outcomes has probability

$$P(A) = \frac{\text{Number of outcomes that make up the event A}}{\text{Total number of outcomes in the sample space}} = \frac{f}{N}.$$

Example

If we toss a coin, the sample space has **two equally likely** outcomes, H and T, only **one** of which is heads. Thus, by the f/N Rule, the **probability** of **heads** is

$$P\left(H\right) = \frac{1}{2},$$

which is also the long-run proportion of times the coin will land heads.

Example

A jar contains **4 red** jelly beans, and **6 yellow** ones. You'll close your eyes and grab a jelly bean. Let

A = The event that you get a **red** jelly bean.

Then by the f/N Rule,

$$P(A) = \frac{4}{10} = \mathbf{0.4}.$$

Example

In the Fall 2016 semester, there were 19,800 students at MSU Denver, of whom 7,812 were students **of color.**

Suppose a student is randomly selected (so that each student is **equally likely** to be selected). Let

A = The event that the selected student is **of color**

Then by the f/N Rule,

$$P(A) = \frac{7,812}{19,800} = \mathbf{0.395}.$$

Basic Properties of Probabilities

- Probabilities satisfy the following properties.

Properties that All Probabilities Satisfy:

1. For any event A, $0 \le P(A) \le 1$.
2. If an event A cannot occur, then $P(A) = 0$.
3. If an event A is certain to occur, then $P(A) = 1$.

Relationships Among Events

- Suppose A and B are two events. Here are **new events** involving A and B:

Relationships Among Events

- Suppose A and B are two events. Here are **new events** involving A and B:
  - The event ***(not A)*** is the event that ”A doesn’t occur.” (Not A) is also called the ***complement*** of A.

Relationships Among Events

- Suppose A and B are two events. Here are **new events** involving A and B:
  - The event ***(not A)*** is the event that "A doesn't occur." (Not A) is also called the ***complement*** of A.

    (Not A) consists of all the outcomes that are **not** in A.

Relationships Among Events

- Suppose A and B are two events. Here are **new events** involving A and B:
  - The event ***(not A)*** is the event that ”A doesn’t occur.” (Not A) is also called the ***complement*** of A.

    (Not A) consists of all the outcomes that are **not** in A.
  - The event ***(A & B)*** is the event that ”A occurs and B occurs.”

Relationships Among Events

- Suppose A and B are two events. Here are **new events** involving A and B:
  - The event ***(not A)*** is the event that "A doesn't occur." (Not A) is also called the ***complement*** of A.

    (Not A) consists of all the outcomes that are **not** in A.
  - The event ***(A & B)*** is the event that "A occurs and B occurs."

    (A & B) consists of the outcomes that are common to both A **and** B.

Relationships Among Events

- Suppose A and B are two events. Here are **new events** involving A and B:
    - The event ***(not A)*** is the event that "A doesn't occur." (Not A) is also called the ***complement*** of A.

      (Not A) consists of all the outcomes that are **not** in A.
    - The event ***(A & B)*** is the event that "A occurs and B occurs."

      (A & B) consists of the outcomes that are common to both A **and** B.
    - The event ***(A or B)*** is the event that "A occurs or B occurs (or both occur)."

Relationships Among Events

- Suppose A and B are two events. Here are **new events** involving A and B:
  - The event ***(not A)*** is the event that "A doesn't occur." (Not A) is also called the ***complement*** of A.

    (Not A) consists of all the outcomes that are **not** in A.
  - The event ***(A & B)*** is the event that "A occurs and B occurs."

    (A & B) consists of the outcomes that are common to both A **and** B.
  - The event ***(A or B)*** is the event that "A occurs or B occurs (or both occur)."

    (A or B) consists of the outcomes that are either in A **or** in B or both.

Example

Roll a die. Let

A = The event that the die lands on an **even** number
= {⚁, ⚃, ⚅}

B = The event that the die lands on a number **4 or higher**
= {⚃, ⚄, ⚅}

Example

Roll a die. Let

A = The event that the die lands on an **even** number
= {⚁, ⚃, ⚅}

B = The event that the die lands on a number **4 or higher**
= {⚃, ⚄, ⚅}

Then

(Not A) = The event that the die lands on an **odd** number
= {⚀, ⚂, ⚄}

A = The event that the die lands on an **even** number
= {⚁, ⚃, ⚅}

B = The event that the die lands on a number **4 or higher**
= {⚃, ⚄, ⚅}

A = The event that the die lands on an **even** number

= {⚁, ⚃, ⚅}

B = The event that the die lands on a number **4 or higher**

= {⚃, ⚄, ⚅}

Also,

(A & B) = The event that the die lands on an **even** number **and** it's **4 or higher**

= {⚃, ⚅}

A = The event that the die lands on an **even** number = {⚁, ⚃, ⚅}

B = The event that the die lands on a number **4 or higher** = {⚃, ⚄, ⚅}

A = The event that the die lands on an **even** number
= {⚁, ⚃, ⚅}

B = The event that the die lands on a number **4 or higher**
= {⚃, ⚄, ⚅}

Finally,

(A *or* B) = The event that die lands on an **even** number **or** a number **4 or higher**
= {⚁, ⚃, ⚄, ⚅}

Mutually Exclusive Events

- Two events A and B are ***mutually exclusive*** if **they can't *both* occur** (i.e. if they *share no outcomes in common*).

Example

Randomly select a student from a university. Let

$$\begin{aligned} A &= \text{The event that the selected student is } \textbf{male} \\ B &= \text{The event that the selected student is } \textbf{pregnant} \end{aligned}$$

Example

Randomly select a student from a university. Let

A = The event that the selected student is **male**
B = The event that the selected student is **pregnant**

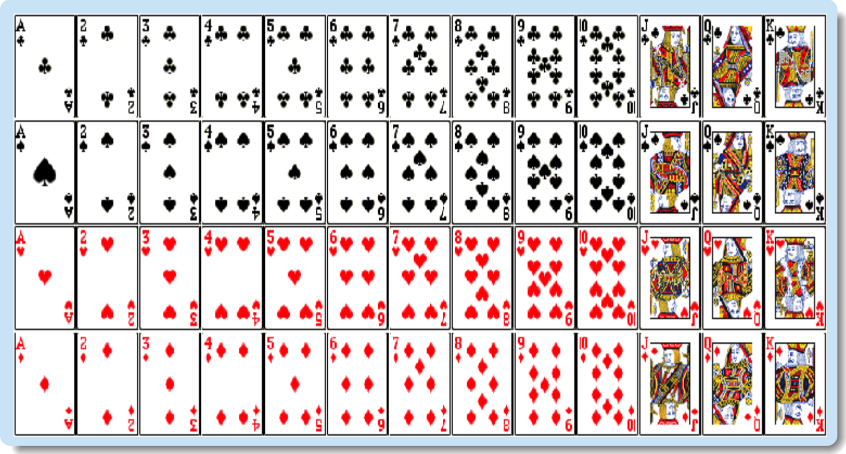
Then A and B are **mutually exclusive** because the selected student *cannot* be both male and pregnant, i.e. there are no students who belong to both the male group *and* the pregnant group.

Example

Recall that in a standard deck of **52 cards**,

Each card is one of **13 ranks** (Ace, 2, 3, ..., Queen, King),

Each card also belongs to one of **four suits** (Clubs $\clubsuit$, Spades $\spadesuit$, Hearts $\heartsuit$, and Diamonds $\diamondsuit$).



A card is to be randomly selected from a deck. Let

$$A = \text{The event that the selected card is an } \textbf{Ace}$$
$$B = \text{The event that the selected card is a } \textbf{Spade}$$

A card is to be randomly selected from a deck. Let

$$A = \text{The event that the selected card is an } \textbf{Ace}$$
$$B = \text{The event that the selected card is a } \textbf{Spade}$$

Then A and B are ***not*** **mutually exclusive** because the selected card *can* be both an Ace *and* a Spade, i.e. there's a card that belongs to both the Ace group *and* the Spade group.

Venn Diagrams

- A ***Venn diagram*** is a graphical depiction of the **sample space** as a **rectangle**, with **events** depicted as **circles** within the rectangle.

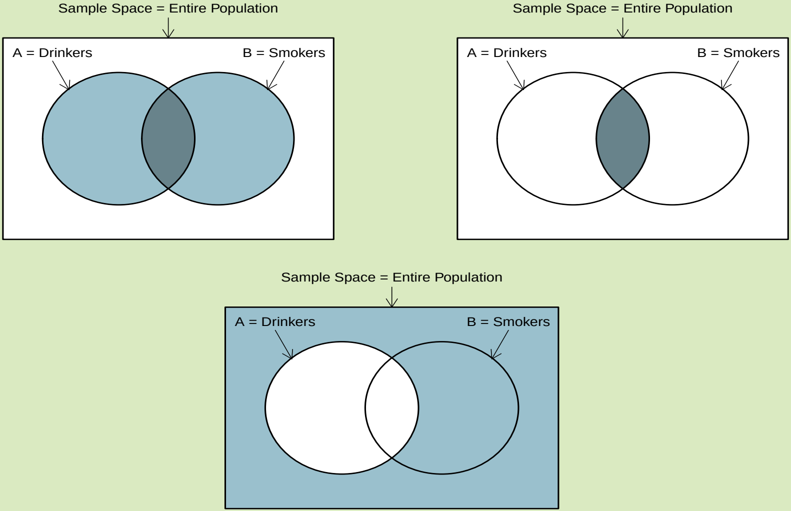
Exercise

In a certain population, some of the people **drink** alcohol and some **smoke** cigarettes.

Consider randomly selecting an individual from this population. Let

A = The event that the selected person **drinks**
B = The event that the selected person **smokes**

a) **Describe** the event **(not A)** in words. Which **Venn diagram** on the next slide depicts this event as its shaded region?

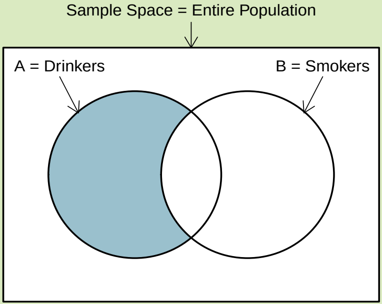

Sample Space = Entire Population
A = Drinkers
B = Smokers
Sample Space = Entire Population
A = Drinkers
B = Smokers
Sample Space = Entire Population
A = Drinkers
B = Smokers

b) **Describe** the event **(A & B)** in words. Which **Venn diagram** on the previous slide depicts this event as its shaded region?

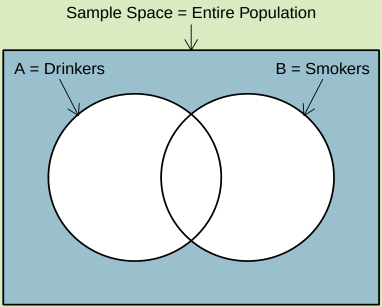
b) **Describe** the event **(A & B)** in words. Which **Venn diagram** on the previous slide depicts this event as its shaded region?

c) **Describe** the event **(A or B)** in words. Which **Venn diagram** on the previous slide depicts this event as its shaded region?

d) **Describe** in words the event that's shaded in the **Venn diagram** below.



e) **Describe** in words the event that's shaded in the **Venn diagram** below.



Exercise

Consider randomly select a student from a university. Let

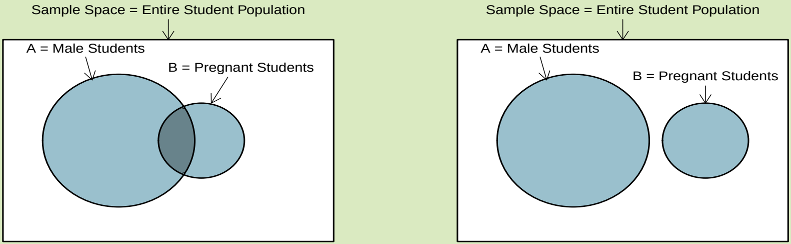
$$\begin{aligned} A &= \text{The event that the selected student is } \textbf{male} \\ B &= \text{The event that the selected student is } \textbf{pregnant} \end{aligned}$$

Exercise

Consider randomly select a student from a university. Let

A = The event that the selected student is **male**
B = The event that the selected student is **pregnant**

a) Which of the two **Venn diagrams** below most accurately depicts these **mutually exclusive** events?



b) In general, if two events A and B are **mutually exclusive**, how should they be depicted in a **Venn diagram**?

Some Probability Rules (4.3)

Complementation Rule and Special Addition Rule

- The following **probability rules** will help us calculate probabilities.

Some Probability Rules (4.3)

Complementation Rule and Special Addition Rule

- The following **probability rules** will help us calculate probabilities.

  1. The ***Complementation Rule*** says:

Complementation Rule: If A is **any** event, then

$$P(\textit{not } A) = 1 - P(A)$$

and also

$$P(A) = 1 - P(\textit{not } A)$$

- (cont'd)

2. The ***Special Addition Rule*** says:

Special Addition Rule: If A and B are **mutually exclusive** events, then

$$P(A \text{ or } B) = P(A) + P(B)$$

Exercise

Suppose a certain university has **1,000** students, of whom **475** are males and **100** are pregnant.

Consider randomly selecting an individual from this university, and let

$$A = \text{The event that the selected person is } \textbf{male}$$
$$B = \text{The event that the selected person is } \textbf{pregnant}$$

Exercise

Suppose a certain university has **1,000** students, of whom **475** are males and **100** are pregnant.

Consider randomly selecting an individual from this university, and let

A = The event that the selected person is **male**
B = The event that the selected person is **pregnant**

a) Use the **Complementation Rule** to find the probability that the selected student is ***not* male**.

Exercise

Suppose a certain university has **1,000** students, of whom **475** are males and **100** are pregnant.

Consider randomly selecting an individual from this university, and let

A = The event that the selected person is **male**
B = The event that the selected person is **pregnant**

a) Use the **Complementation Rule** to find the probability that the selected student is ***not*** **male.**

b) Use the **Special Addition Rule** to find the probability that the selected student is **male *or* pregnant**. (The rule is applicable because A and B are **mutually exclusive**.)

Exercise

Suppose that in a population of **1,000** people, **300 drink** alcohol and **100 smoke** cigarettes.

Consider randomly selecting an individual from this population, and let

$$\begin{aligned} A &= \text{The event that the selected person drinks} \\ B &= \text{The event that the selected person smokes} \end{aligned}$$

Exercise

Suppose that in a population of **1,000** people, **300 drink** alcohol and **100 smoke** cigarettes.

Consider randomly selecting an individual from this population, and let

$$\begin{aligned} A &= \text{The event that the selected person drinks} \\ B &= \text{The event that the selected person smokes} \end{aligned}$$

Is the **Special Addition Rule** applicable for finding the **probability** that the selected person **drinks *or* smokes**? **Why or why not**?

The General Addition Rule

- **When two events A and B *aren't* mutually exclusive, we *can't* use the Special Addition Rule** to find P(A *or* B).

The General Addition Rule

- **When two events A and B *aren't* mutually exclusive, we *can't* use the Special Addition Rule** to find P(A *or* B).

  The next example shows what goes wrong.

Example

A library has **100 books.** **Seventy** of the books cover **statistics** and **40** cover **probability**.

Example

A library has **100 books. Seventy** of the books cover **statistics** and **40** cover **probability.**

We want to find the probability that a randomly selected book covers **statistics *or* probability.** Let

A = The event that the selected book covers **statistics**
B = The event that the selected book covers **probability**

Then $P(A) = \frac{\mathbf{70}}{\mathbf{100}}$ and $P(B) = \frac{\mathbf{40}}{\mathbf{100}}$.

Example

A library has **100 books. Seventy** of the books cover **statistics** and **40** cover **probability.**

We want to find the probability that a randomly selected book covers **statistics *or* probability.** Let

A = The event that the selected book covers **statistics**
B = The event that the selected book covers **probability**

Then $P(A) = \frac{70}{100}$ and $P(B) = \frac{40}{100}$.

Is it true that

$$P(A \text{ or } B) = P(A) + P(B)$$ **??????**

No!

No!

P(A) and P(B) add up to $110/100$, but probabilities **can't** be bigger than 1.

No!

P(A) and P(B) add up to $110/100$, but probabilities **can't** be bigger than 1.

What went wrong?

No!

P(A) and P(B) add up to $110/100$, but probabilities **can't** be bigger than 1.

What went wrong?

Well, some of the books cover *both* topics. In other words, **A and B *aren't* mutually exclusive.**

No!

P(A) and P(B) add up to $110/100$, but probabilities **can't** be bigger than 1.

What went wrong?

Well, some of the books cover *both* topics. In other words, **A and B *aren't* mutually exclusive.**

So when we added $70/100$ to $40/100$, we **double-counted** the books that cover *both* topics.

No!

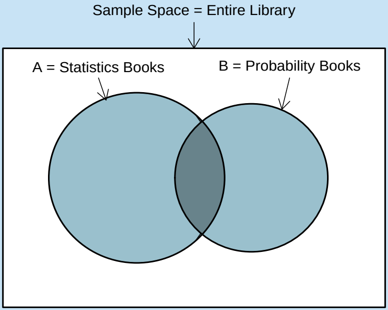
P(A) and P(B) add up to $110/100$, but probabilities **can't** be bigger than 1.

What went wrong?

Well, some of the books cover *both* topics. In other words, **A and B *aren't* mutually exclusive.**

So when we added $70/100$ to $40/100$, we **double-counted** the books that cover *both* topics.

The double-counted books are the ones in the overlap of the circles in the Venn diagram (next slide).


Sample Space = Entire Library
A = Statistics Books
B = Probability Books

- The ***General Addition Rule*** applies even when events A and B **aren't** mutually exclusive.

General Addition Rule: For **any** two events A and B (**not necessarily mutually exclusive**),

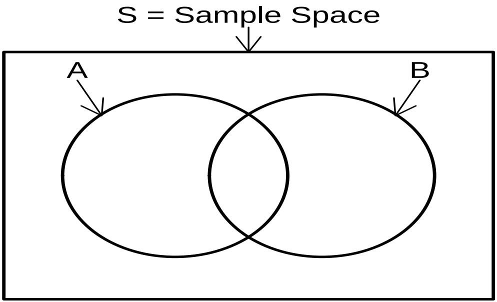
$$P(A \text{ or } B) = P(A) + P(B) - P(A \,\&\, B)$$

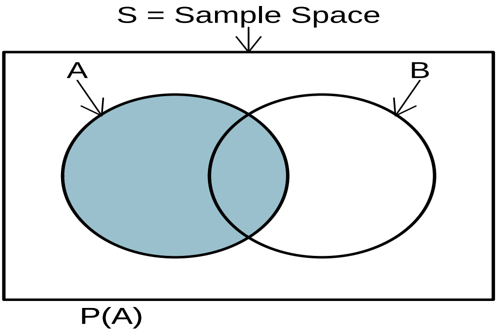
- The ***General Addition Rule*** applies even when events A and B **aren't** mutually exclusive.

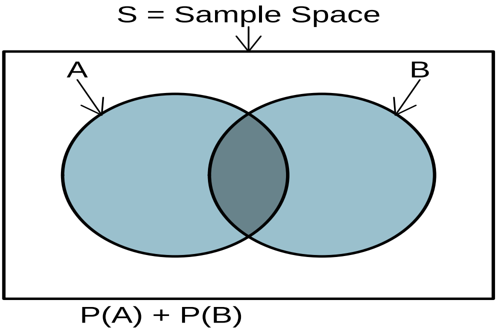
General Addition Rule: For **any** two events A and B (**not necessarily mutually exclusive**),

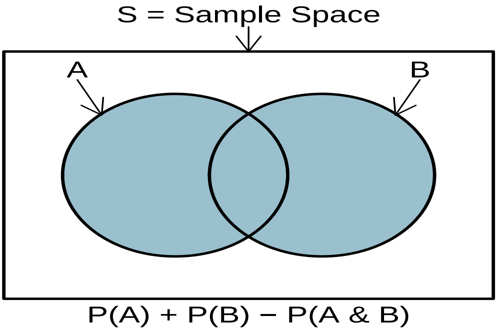
$$P(A \text{ or } B) \;=\; P(A) + P(B) - P(A \;\&\; B)$$

Subtracting P(A & B) corrects for the fact that when A and B *aren't* mutually exclusive, some outcomes are **double-counted** when P(B) and P(A) are added together.


S = Sample Space
A
B


S = Sample Space
A
B
P(A)


S = Sample Space
A
B
P(A) + P(B)


S = Sample Space
A
B
P(A) + P(B) – P(A & B)

Exercise

In the library that has **100** books, **70** of which cover *statistics* and **40** of which cover *probability*, suppose that **30** books cover ***both*** topics.

Exercise

In the library that has **100** books, **70** of which cover *statistics* and **40** of which cover *probability*, suppose that **30** books cover ***both*** topics.

For a randomly selected book, let

$$A = \text{The event that the selected book covers } \textbf{statistics}$$
$$B = \text{The event that the selected book covers } \textbf{probability}$$

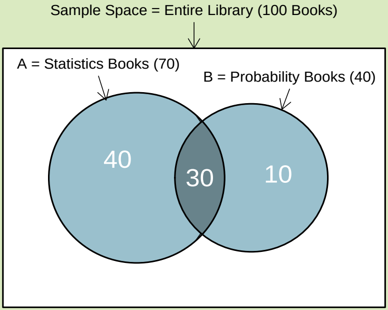
Exercise

In the library that has **100** books, **70** of which cover *statistics* and **40** of which cover *probability*, suppose that **30** books cover ***both*** topics.

For a randomly selected book, let

$$
\begin{aligned}
A &= \text{The event that the selected book covers } \textbf{statistics} \\
B &= \text{The event that the selected book covers } \textbf{probability}
\end{aligned}
$$

Refer to the following Venn diagram.


Sample Space = Entire Library (100 Books)
A = Statistics Books (70)
B = Probability Books (40)
40
30
10

a) Use the **General Addition Rule** to find the probability that a randomly selected book covers **statistics *or* probability**.

a) Use the **General Addition Rule** to find the probability that a randomly selected book covers **statistics *or* probability**.

b) Verify the answer to Part *a* via the Venn diagram.

Exercise

In a population of **1,000 people**, **300** drink alcohol and **100** smoke cigarettes. Suppose also that **80** people have **both** of these bad habits.

For a randomly selected individual, let

$$A = \text{The event that the selected person } \textbf{drinks}$$
$$B = \text{The event that the selected person } \textbf{smokes}$$

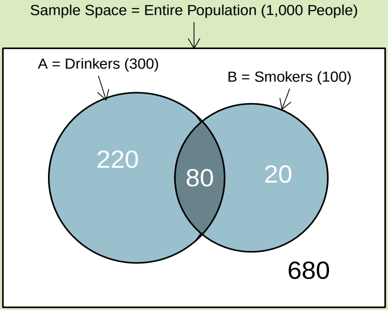
Exercise

In a population of **1,000 people**, **300** drink alcohol and **100** smoke cigarettes. Suppose also that **80** people have **both** of these bad habits.

For a randomly selected individual, let

$$A = \text{The event that the selected person } \textbf{drinks}$$
$$B = \text{The event that the selected person } \textbf{smokes}$$

Refer to the following Venn diagram.


Sample Space = Entire Population (1,000 People)
A = Drinkers (300)
B = Smokers (100)
220
80
20
680

a) Use the **General Addition Rule** to find the probability that the selected person **drinks *or* smokes.**

a) Use the **General Addition Rule** to find the probability that the selected person **drinks *or* smokes.**

b) Verify the answer to Part *a* via the Venn diagram.