

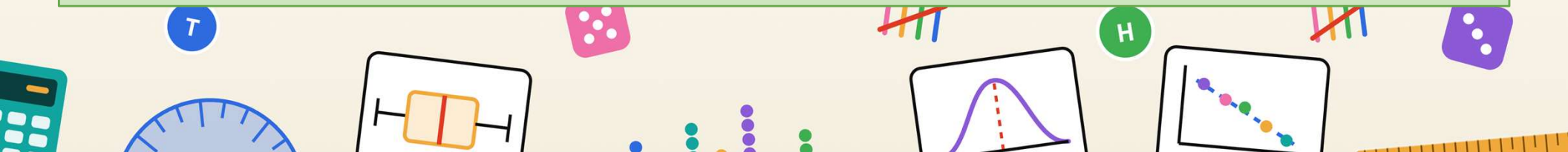
Analyzing Data on Two Categorical Variables

How do you analyze data when a data set involves *two* categorical variables?

Respondent	Environmental club?	Snowmobile use
1	No	Own
2	No	Rent
3	Yes	Never
4	Yes	Rent
5	No	Never
⋮	⋮	⋮

A **two-way table** is a table of counts that summarizes data on the relationship between two categorical variables for some group of individuals.

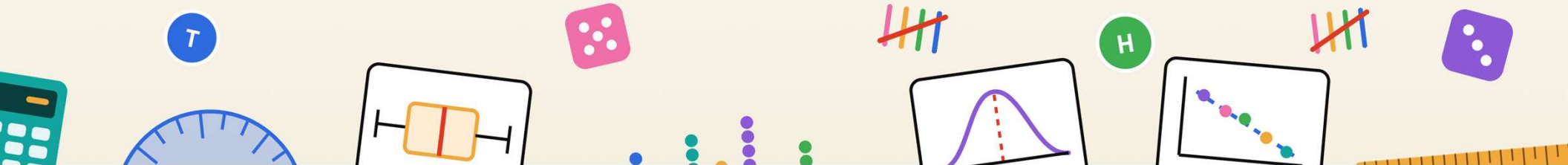
A two-way table also goes by the name **contingency table**. Same table, same counts - the AP exam uses both names.



Analyzing Data on Two Categorical Variables (cont.)

Snowmobile use	Club member: No	Club member: Yes	Total
Never used	445	212	657
Snowmobile renter	497	77	574
Snowmobile owner	279	16	295
Total	1221	305	1526

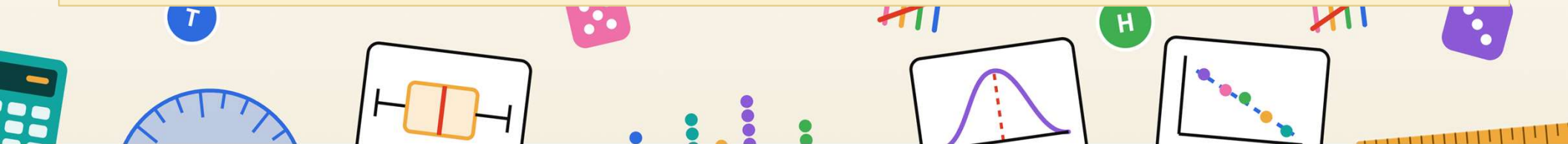
We can include row and column totals. Before computing anything, make sure the table adds up - every path to the grand total must agree:



Analyzing Data on Two Categorical Variables (cont., part 2)

Snowmobile use	Club member: No	Club member: Yes	Total
Never used	445	212	657
Snowmobile renter	497	77	574
Snowmobile owner	279	16	295
Total	1221	305	1526

- Sum of all six cells: $445 + 212 + 497 + 77 + 279 + 16 = 1526$
- Sum of the row totals: $657 + 574 + 295 = 1526$
- Sum of the column totals: $1221 + 305 = 1526$

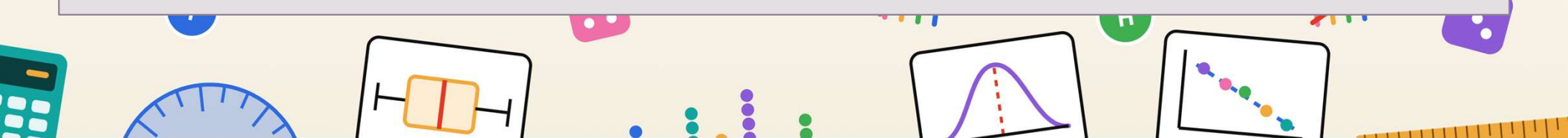


Reading the Margins

A **marginal relative frequency** gives the percent or proportion of individuals that have a specific value for one categorical variable.

Snowmobile use	Club member: No	Club member: Yes	Total
Never used	445	212	657
Snowmobile renter	497	77	574
Snowmobile owner	279	16	295
Total	1221	305	1526

- (a) Find the marginal relative frequencies for environmental club membership.
- (b) Find the marginal relative frequencies for snowmobile use.

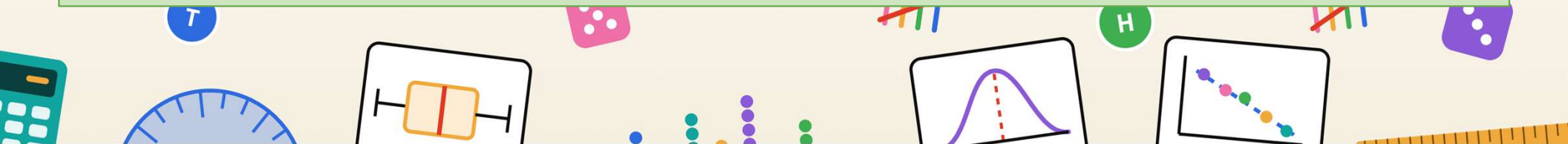


Reading the Margins (cont.)

Snowmobile use	Club member: No	Club member: Yes	Total
Never used	445	212	657
Snowmobile renter	497	77	574
Snowmobile owner	279	16	295
Total	1221	305	1526

(a) No: $1221/1526 = 0.800$ or 80.0%. Yes: $305/1526 = 0.200$ or 20.0%.

(b) Never: $657/1526 = 0.431$ or 43.1%. Rent: $574/1526 = 0.376$ or 37.6%. Own: $295/1526 = 0.193$ or 19.3%.

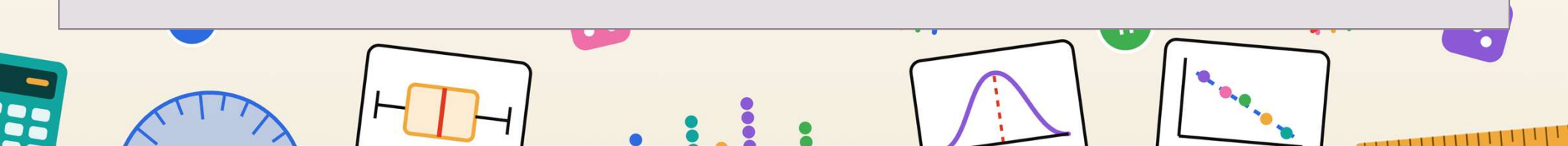


Where Row Meets Column

A **joint relative frequency** gives the percent or proportion of individuals that have a specific value for one categorical variable and a specific value for another categorical variable.

Snowmobile use	Club member: No	Club member: Yes	Total
Never used	445	212	657
Snowmobile renter	497	77	574
Snowmobile owner	279	16	295
Total	1221	305	1526

- (a) What percent of people in the sample are environmental club members *and* own snowmobiles?
- (b) What proportion of people in the sample are not environmental club members *and* never use snowmobiles?

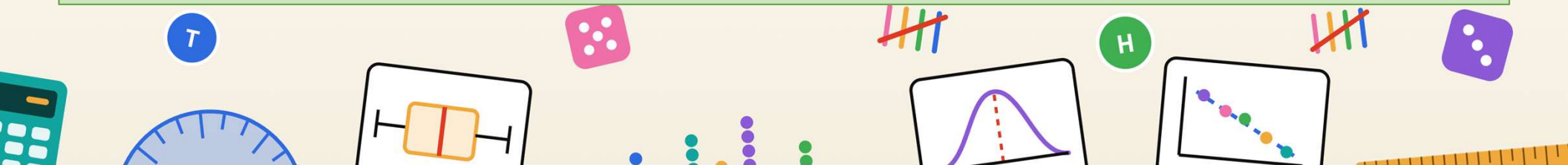


Where Row Meets Column (cont.)

Snowmobile use	Club member: No	Club member: Yes	Total
Never used	445	212	657
Snowmobile renter	497	77	574
Snowmobile owner	279	16	295
Total	1221	305	1526

$$(a) 16/1526 = 0.010 = 1.0\%$$

$$(b) 445/1526 = 0.292$$



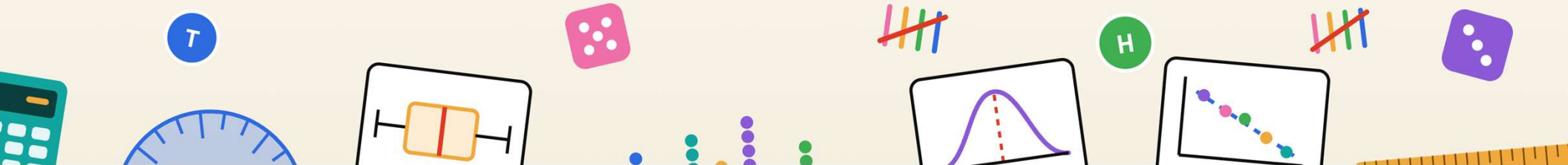
Who Wrote the Hits?

The Beatles have more #1 songs on the Billboard Hot 100 than any other artist or band. Who wrote them? All 205 Beatles songs released in the U.S.:

	George Harrison	John Lennon	Paul McCartney	Ringo Starr	Multiple writers	Total
Reached #1	0	5	12	0	3	20
Did not reach #1	25	69	63	2	26	185
Total	25	74	75	2	29	205

AP[®] Exam Tip

Why "joint"? The joints in your body are where two bones come together. A joint relative frequency is found where a row and a column come together.



Who Wrote the Hits? (cont.)

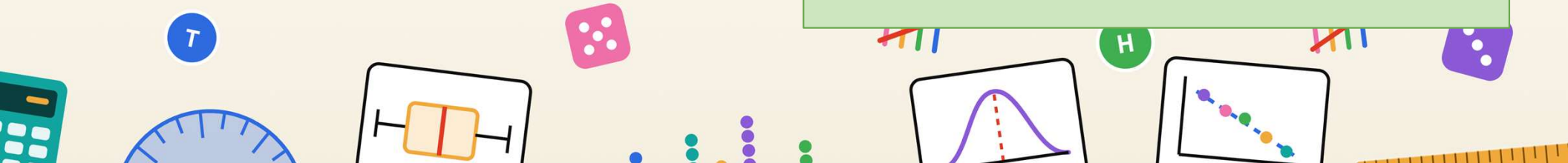
	George Harrison	John Lennon	Paul McCartney	Ringo Starr	Multiple writers	Total
Reached #1	0	5	12	0	3	20
Did not reach #1	25	69	63	2	26	185
Total	25	74	75	2	29	205

(a) What percentage of Beatles songs reached number one?

(b) What proportion of Beatles songs were written by Paul McCartney alone and reached number one?

(a) Marginal relative frequency: $20/205 = 0.098$, so about 9.8% of all Beatles songs reached #1.

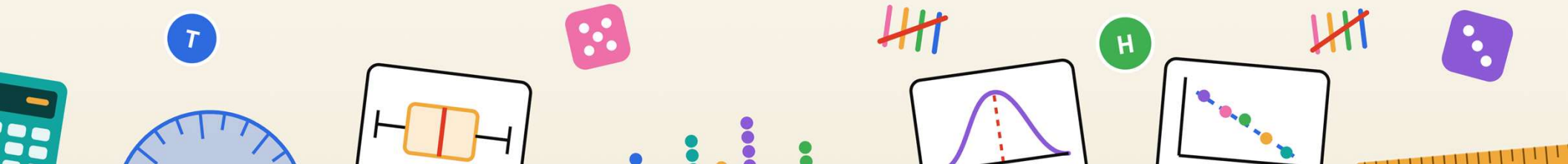
(b) Joint relative frequency: $12/205 = 0.059$, so about 5.9% were McCartney-alone songs that reached #1.



The Denominator Follows the “Of”

Both questions are **conditional relative frequencies** - but they condition in opposite directions. (c) conditions on a *writer column*; (d) conditions on a *#1 row*.

	George Harrison	John Lennon	Paul McCartney	Ringo Starr	Multiple writers	Total
Reached #1	0	5	12	0	3	20
Did not reach #1	25	69	63	2	26	185
Total	25	74	75	2	29	205



The Denominator Follows the “Of” (cont.)

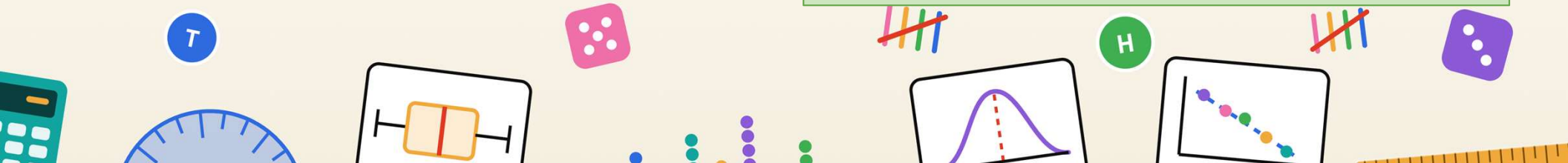
	George Harrison	John Lennon	Paul McCartney	Ringo Starr	Multiple writers	Total
Reached #1	0	5	12	0	3	20
Did not reach #1	25	69	63	2	26	185
Total	25	74	75	2	29	205

(c) What percentage of songs written by Paul McCartney alone reached number one?

(d) What proportion of number one songs were written by Paul McCartney alone?

(c) Condition: McCartney-alone songs. $12/75 = 16\%$ of *McCartney's songs* reached #1.

(d) Condition: #1 songs. $12/20 = 60\%$ of *the #1 songs* were McCartney's alone.

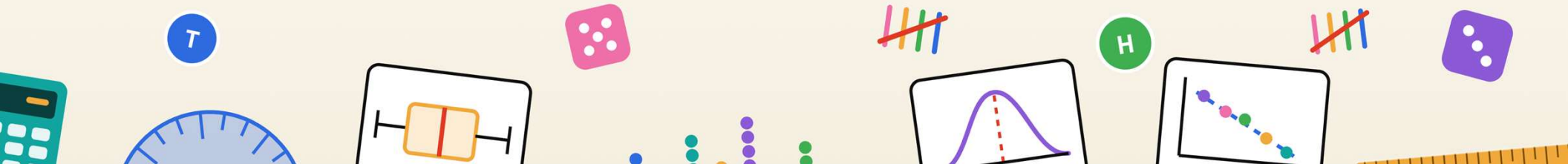


The Denominator Follows the “Of” (cont.)

	George Harrison	John Lennon	Paul McCartney	Ringo Starr	Multiple writers	Total
Reached #1	0	5	12	0	3	20
Did not reach #1	25	69	63	2	26	185
Total	25	74	75	2	29	205

AP® Exam Tip

When translating words into proportions, the value in the denominator is found next to the word "of."
Percentage OF McCartney songs? Divide by 75. Proportion OF number one songs? Divide by 20.



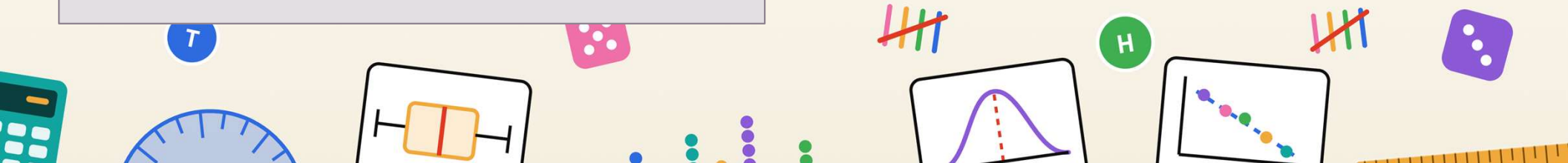
Adding a Condition

Marginal and joint relative frequencies do not tell us much about the *relationship* between environmental club membership and snowmobile use for the people in the sample.

A **conditional relative frequency** gives the percent or proportion of individuals that have a specific value for one categorical variable among individuals who share the same value of another categorical variable (the condition).

What percent of environmental club members in the sample are snowmobile owners?

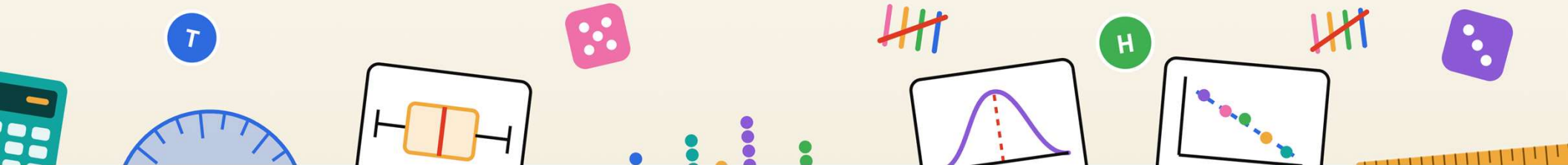
Condition: the 305 club members. $16/305 = 0.052 = 5.2\%$



The Whole Column at Once

The distribution of snowmobile use among environmental club members is called the **conditional distribution** of snowmobile use among environmental club members.

Snowmobile use	Club member: No	Club member: Yes	Total
Never used	445	212	657
Snowmobile renter	497	77	574
Snowmobile owner	279	16	295
Total	1221	305	1526



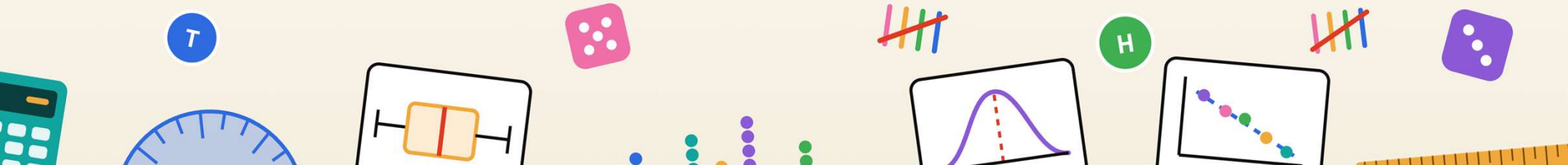
The Whole Column at Once (cont., part 2)

(a) Find the conditional distribution of snowmobile use among environmental club members.

(b) We can find the distribution of snowmobile use among the survey respondents who are not environmental club members in a similar way.

(a) Never: $212/305 = 69.5\%$. Rent: $77/305 = 25.2\%$. Own: $16/305 = 5.2\%$.

(b) Never: $445/1221 = 36.4\%$. Rent: $497/1221 = 40.7\%$. Own: $279/1221 = 22.9\%$.

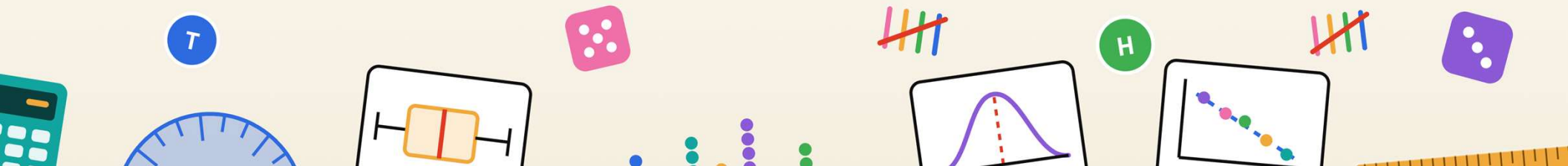


Percents, Not Counts

AP® Exam Tip

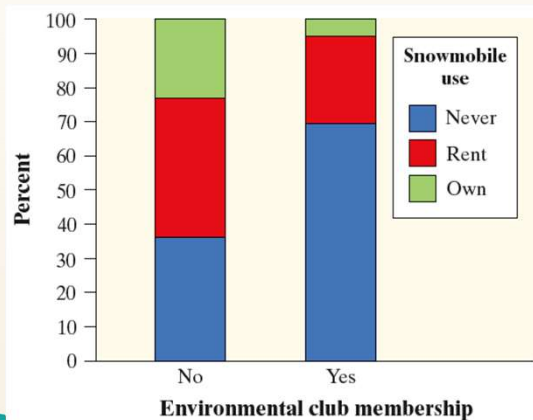
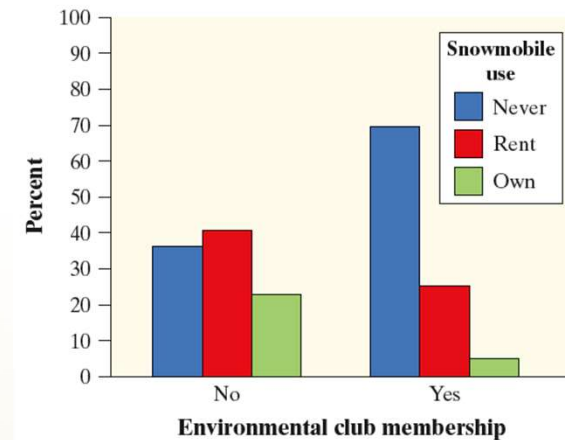
When comparing groups of different sizes, be sure to use relative frequencies (percents or proportions) instead of frequencies (counts) when analyzing categorical data.

Make sure to avoid statements like "More club members never use snowmobiles" when you mean "A greater percentage of club members never use snowmobiles."



Side by Side or Stacked?

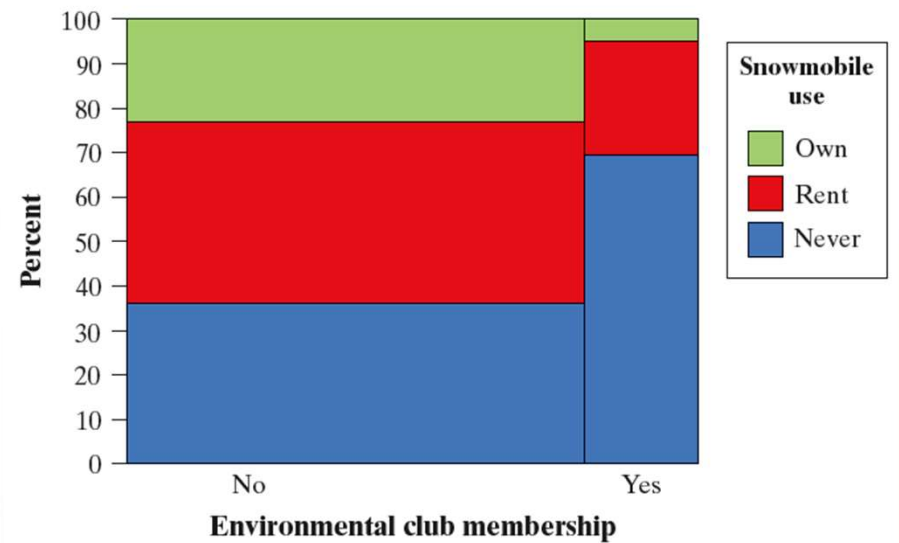
A **side-by-side bar graph** displays the distribution of a categorical variable for each value of another categorical variable. The bars are grouped together based on the values of one of the categorical variables and placed side by side.



A **segmented bar graph** displays the distribution of a categorical variable as segments of a rectangle, with the area of each segment proportional to the percent of individuals in the corresponding category.

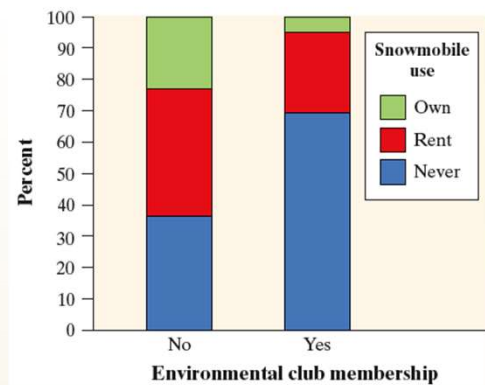
One More Graph: The Mosaic Plot

A **mosaic plot** is a modified segmented bar graph in which the width of each rectangle is proportional to the number of individuals in the corresponding category.

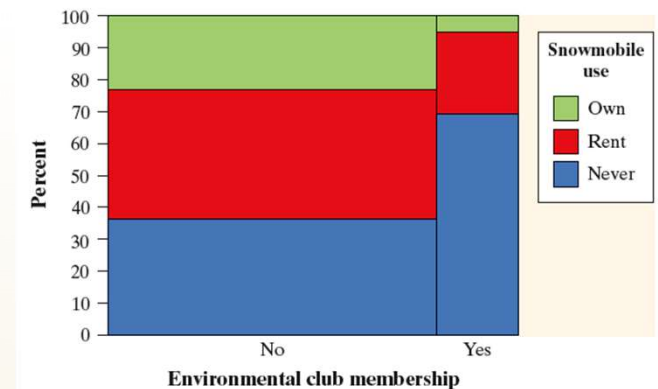


What the Width Is Telling You

Same data, same segment heights - one difference. In a **mosaic plot**, the *width* of each bar is proportional to the number of individuals in that group.



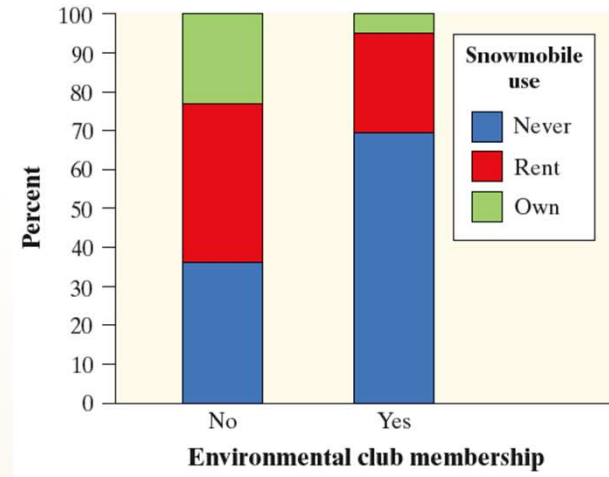
Which graph shows that far more respondents are NOT environmental club members?



The mosaic plot. The No bar is much wider - 1221 of the 1526 respondents are not club members. The segmented bar graph hides the group sizes completely.

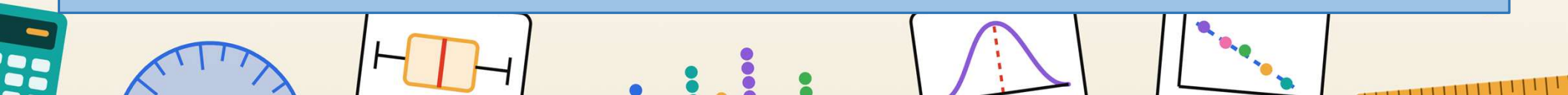
When Knowing One Predicts the Other

There is an **association** between two variables if knowing the value of one variable helps us predict the value of the other.



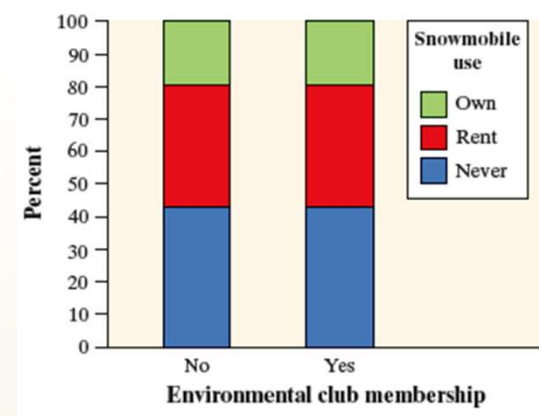
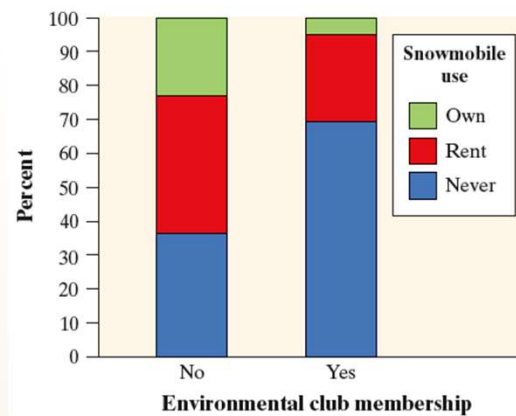
If knowing the value of one variable does not help us predict the value of the other, then there is **no association** between the variables.

CAUTION: Association does not necessarily imply causation!



What NO Association Looks Like

If there were **no association**, knowing whether someone is a club member would tell you *nothing* about their snowmobile use.



The real data

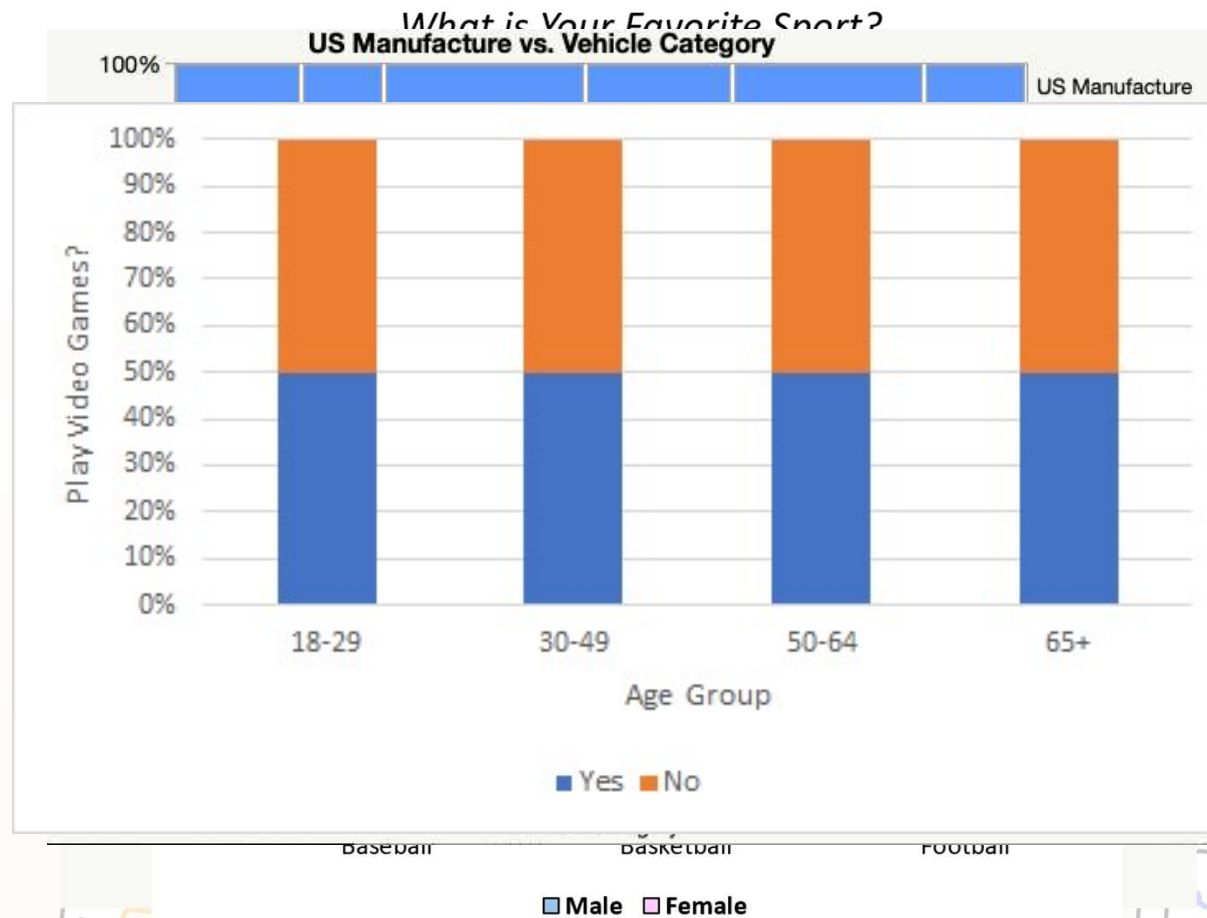
Segment heights DIFFER between the No and Yes groups - there is an association.

A no-association world

Every group's segments sit at the SAME heights - the distribution of snowmobile use is identical in each group.

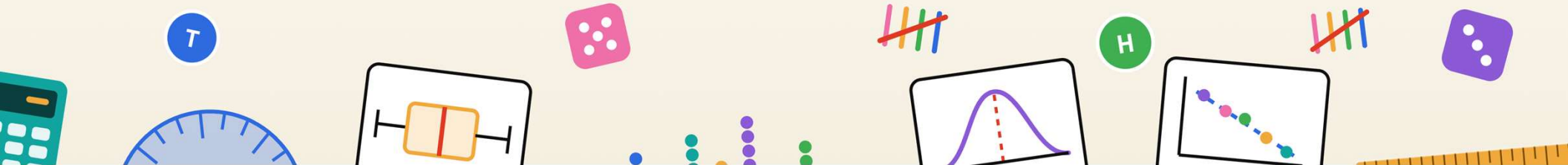


Practice Problems – Associated or Not?



Practice Problem - Titanic

	Survived	Did Not Survive	Total
First Class Passengers	201	123	324
Second Class Passengers	118	166	284
Third Class Passengers	181	528	709
Total Passengers	500	817	1317



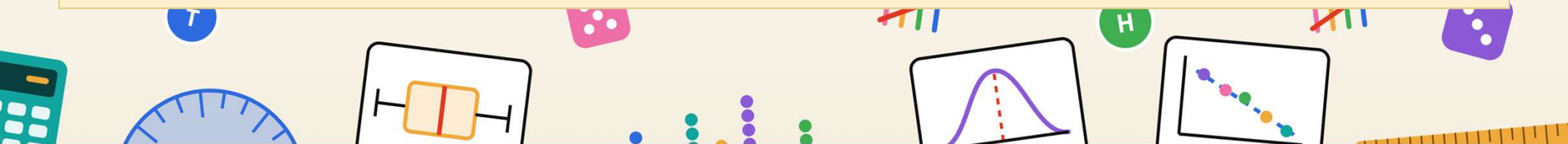
Practice Problem - Titanic

	Survived	Did Not Survived	Total
First Class Passengers	201	123	324
Second Class Passengers	118	166	284
Third Class Passengers	181	528	709
Total Passengers	500	817	1317

1. Calculate the Marginal Relative Frequency for 3rd Class.
2. Calculate the Joint Relative Frequency for 2nd Class Survival.
3. Calculate the Conditional Relative Frequency for Survival given passengers are members of the 1st Class.
4. Is there an association between class and survival?

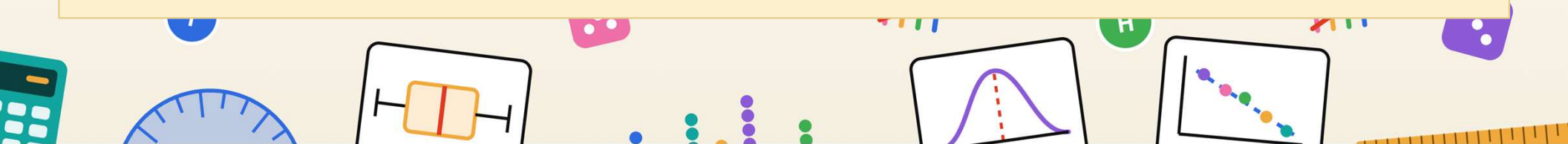
Probability Preview

- When the weather forecast says there's a 60% chance of rain, what does that actually mean?
- How can I know if a coin is "fair"?
- Why do casinos always make money despite paying out large prizes?
- How does the lottery benefit a state's economy?
- If everyone has exactly one soul-mate, what is the chance of never meeting them?
- When a doctor says a medical test is 95% accurate, what does that mean for an individual patient? How should probability affect the way we interpret medical test results?



Probability Preview: Coming Attractions

- If there's a 15% chance of a traffic delay each morning, what's the probability of being delayed on 2 out of 5 days in a week?
- If a student applies to three colleges, each with an independent 30% acceptance rate, what is the probability that they'll get into at least one?
- In a game where there's a 1 in 5 chance of winning a prize, how many times should a player expect to play to win once? How would the answer change if there's a 1 in 10 chance?
- What is the probability of getting at least 10 likes on a post if each friend has a 60% chance of liking it? How does this probability change with a different number of friends?



The Idea of Probability

The **law of large numbers** says that if we observe more and more repetitions of any chance process, the proportion of times that a specific outcome occurs gets closer and closer to its probability.

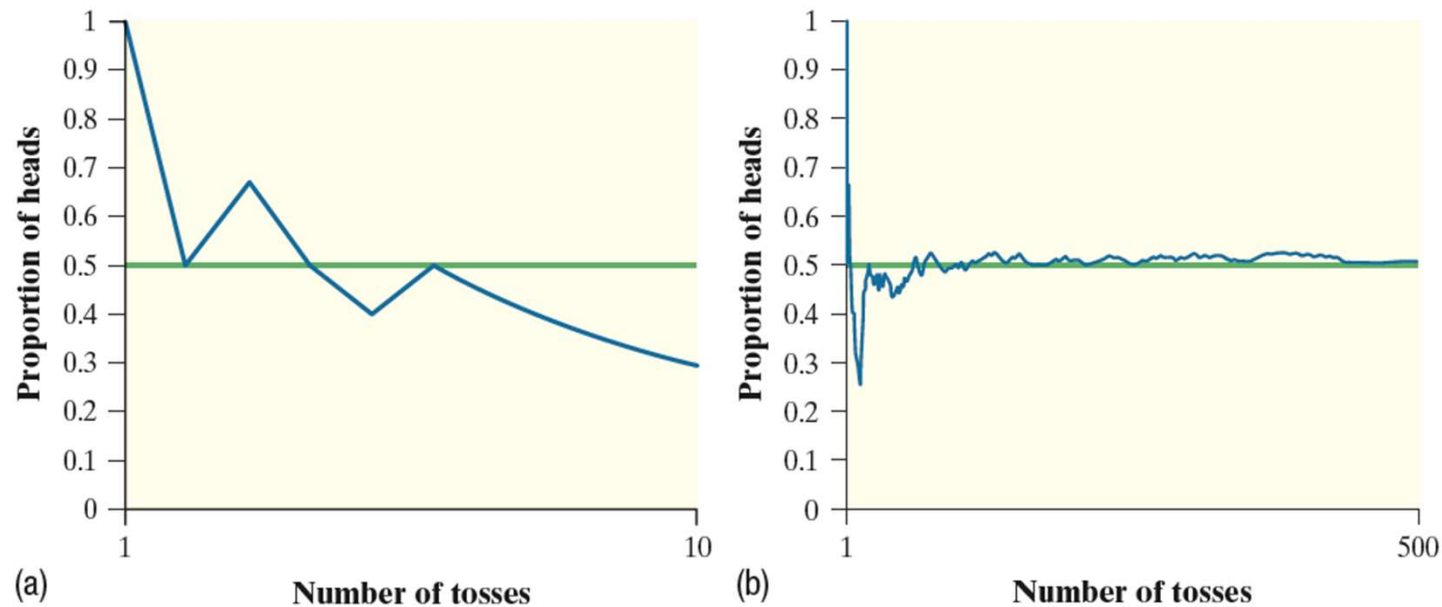


FIGURE 5.1 (a) The proportion of heads in the first 10 tosses of a coin. (b) The proportion of heads in the first 500 tosses of a coin.

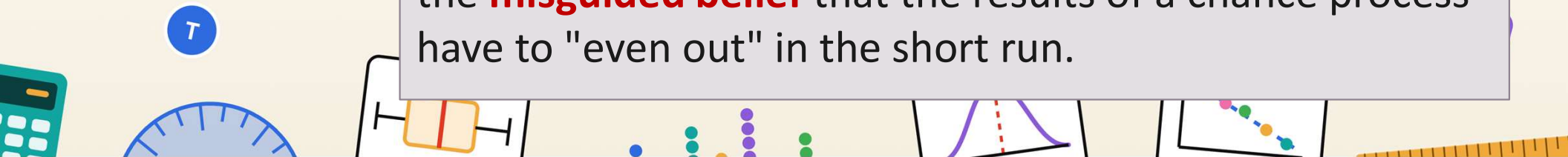
Understanding Randomness

The idea of probability is that randomness is predictable in the long run.

Unfortunately, our intuition about chance behavior tries to tell us that randomness should also be predictable in the short run.

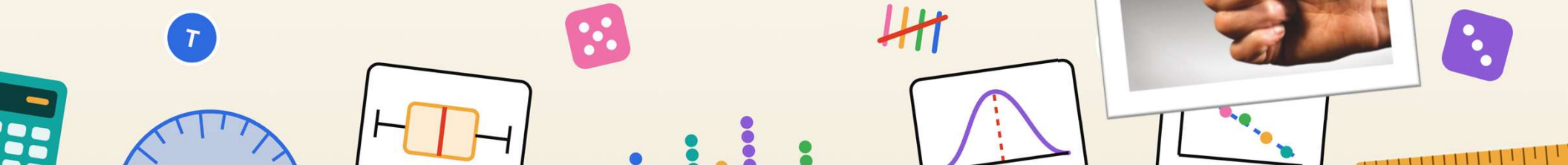


Some people use the phrase *law of averages* to refer to the **misguided belief** that the results of a chance process have to "even out" in the short run.



Understanding Check

There are two games involving flipping a fair coin. In the first game, you win a prize if you can throw between 45% and 55% heads. In the second game, you win if you can throw more than 60% heads. For each game, would you rather flip 20 times or 200 times?



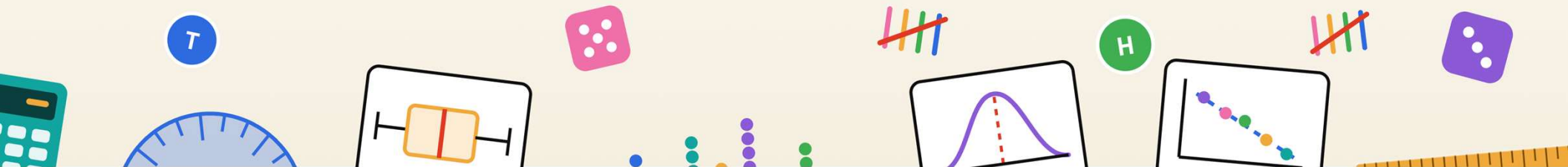
Outcomes and Events

A random process generates outcomes that are determined purely by chance.
A **trial** is one repetition of a random process.

An **outcome** is the result of one trial of a random process.

An **event** is any collection of outcomes from some chance process.

Random process $\rightarrow$ trial $\rightarrow$ outcome $\rightarrow$ event. One flip of a coin is a trial; landing heads is an outcome; "getting at least one head in three flips" is an event.



Where Does a Probability Live?

Every probability is a number between **0** and **1**. An event that never occurs has probability 0; an event that occurs on every trial has probability 1.

(a) Place it: rolling a 7 on one standard die

(b) Place it: a fair coin lands heads

(c) Place it: the sun rises tomorrow

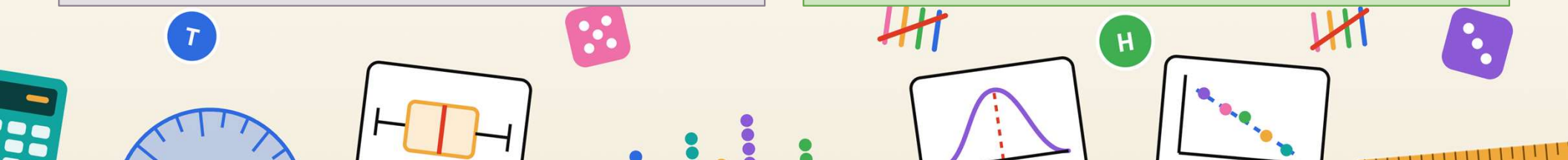
(d) Place it: rolling doubles with two dice

(a) At 0 — impossible

(b) At 0.5 — a 50-50 chance

(c) At (essentially) 1 — happens every trial

(d) $6/36 \approx 0.17$ — unlikely, but far from impossible



Interpret This: The Future Is Bright

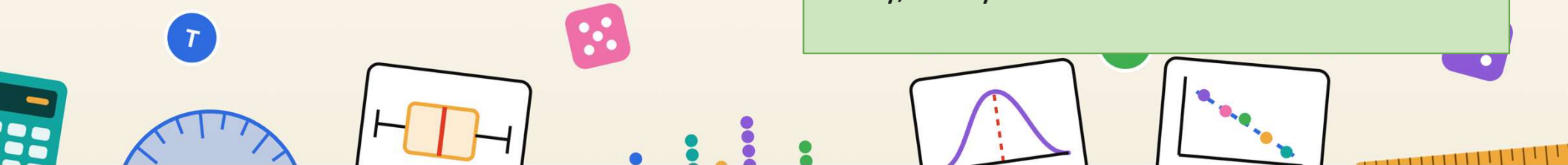
Gallup research: 76% of U.S. Gen Z members believe they have a great future ahead of them. So the probability that a randomly selected U.S. Gen Zer believes this is 0.76.

(a) Interpret this probability.

(b) If a researcher randomly selects 100 U.S. Gen Zers, will exactly 76 of them believe they have a great future ahead of them? Explain.

(a) If you take a very large random sample of U.S. Gen Z members, approximately 76% of them will believe that they have a great future ahead of them.

(b) Probably not. With only 100 selected, the number who believe it is likely to differ from 76. Probability describes what happens in many, many trials — far more than 100.



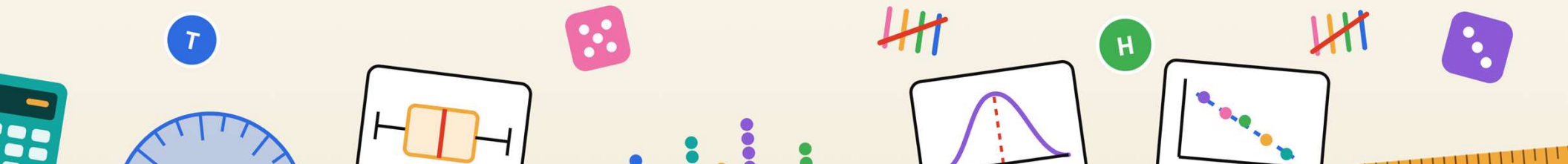
Quick Check: Short-Run Superstitions

(a) I just flipped a fair coin 6 times and got tails every time! Is heads more likely on the next flip?

(b) Which is more likely in 6 flips of a fair coin: TTTTTT or THHTTH?

(a) NO! Still 0.50. Heads appears half the time in the long run, but future outcomes do not have to make up for an imbalance like six straight tails.

(b) Neither — if you are looking for that specific arrangement of heads and tails, they are equally likely.

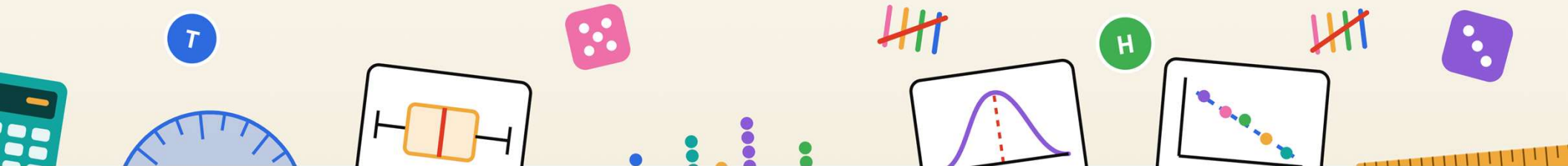


Simulation

Simulation is the imitation of chance behavior, based on a model that accurately reflects the situation.

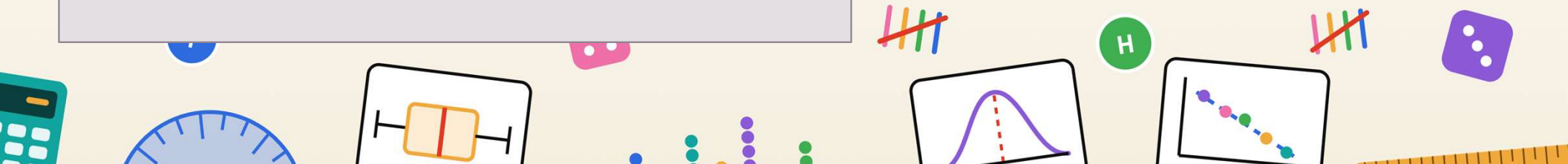
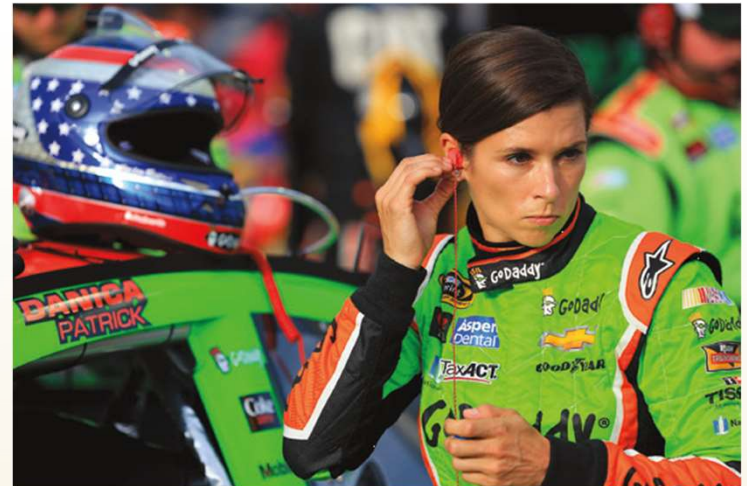
The Simulation Process:

1. Describe how to use a chance device to imitate one trial (repetition) of the simulation. Tell what you will record at the end of each trial.
2. Perform many trials of the simulation.
3. Use the results of your simulation to estimate the probability and answer the question of interest.

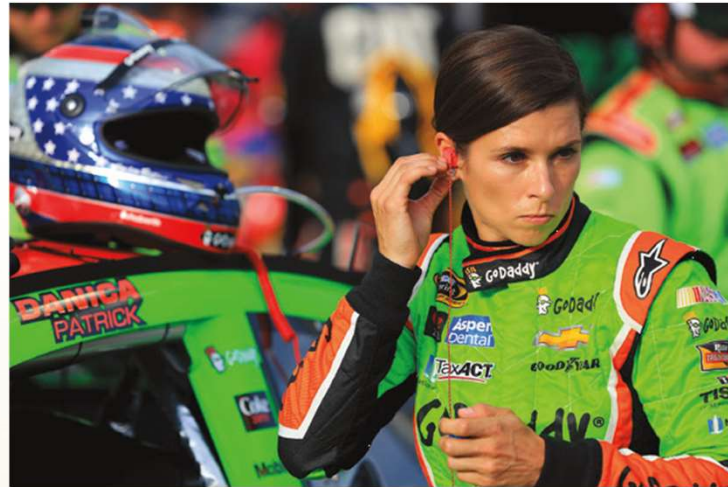


The 23-Box Problem

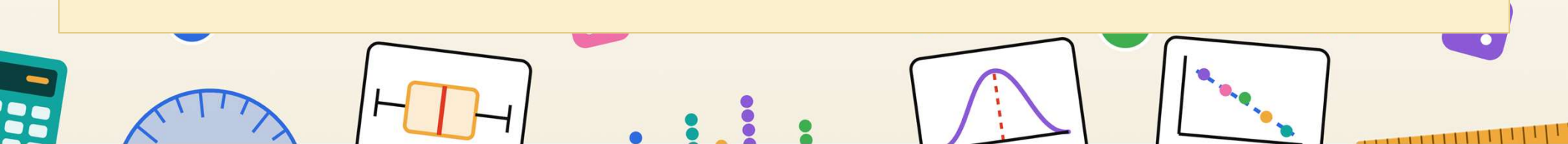
To increase sales, a breakfast cereal company offers a NASCAR promotion: each box contains a collectible card featuring one of 5 drivers — Joey Logano, Kevin Harvick, Chase Elliott, Danica Patrick, or Jimmie Johnson. The company claims each of the 5 cards is equally likely to appear in any box. A NASCAR fan keeps buying boxes until she has all 5 drivers' cards — and is surprised when it takes her **23 boxes**. Does this outcome provide convincing evidence that the 5 cards are not equally likely?



Let's Simulate!!!



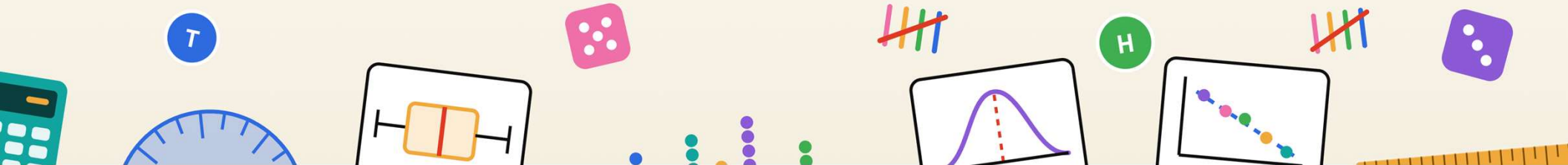
The plan: perform a simulation to estimate the probability that it takes 23 or more boxes to get a full set of 5 NASCAR collectible cards.



The 23-Box Problem: Describing One Trial

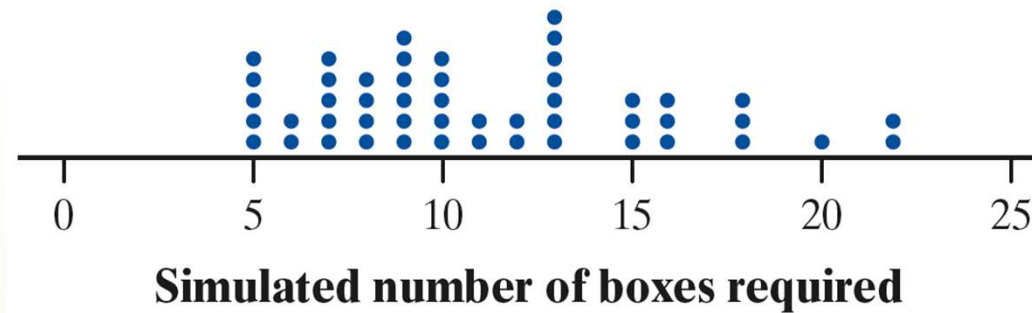
(a) Describe how to use a random number generator to perform one trial of the simulation.

(a) Let 1 = Joey Logano, 2 = Kevin Harvick, 3 = Chase Elliott, 4 = Danica Patrick, 5 = Jimmie Johnson. Generate a random integer from 1 to 5 to simulate buying one box of cereal and looking at which card is inside. Keep generating random integers until all five labels from 1 to 5 appear. Record the number of boxes it takes to get all 5 cards.



The 23-Box Problem: Convincing Evidence?

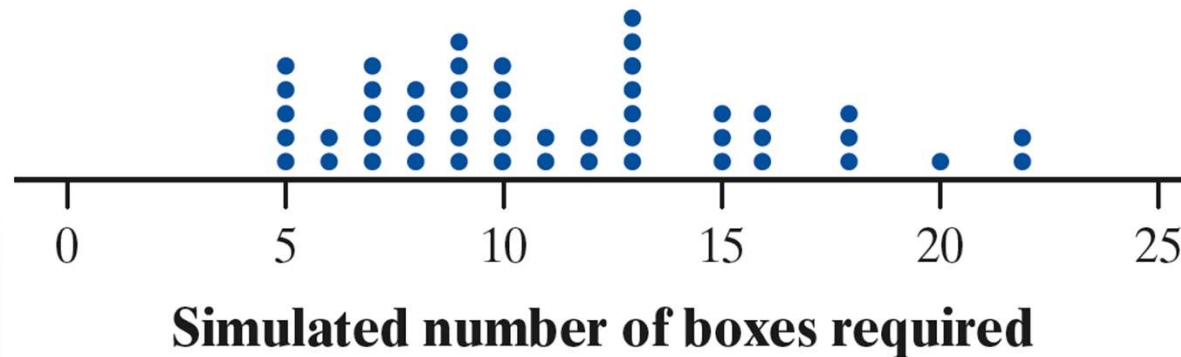
The dotplot shows the number of cereal boxes it took to get all 5 drivers' cards in 50 trials.



(b) Explain what the dot at 20 represents.

(c) Use the results of the simulation to estimate the probability that it will take 23 or more boxes to get a full set of cards. Does this outcome provide convincing evidence that the 5 cards are not equally likely?

The 23-Box Problem: Convincing Evidence? — Solutions



(b) One trial where it took 20 boxes to get all 5 drivers' cards.

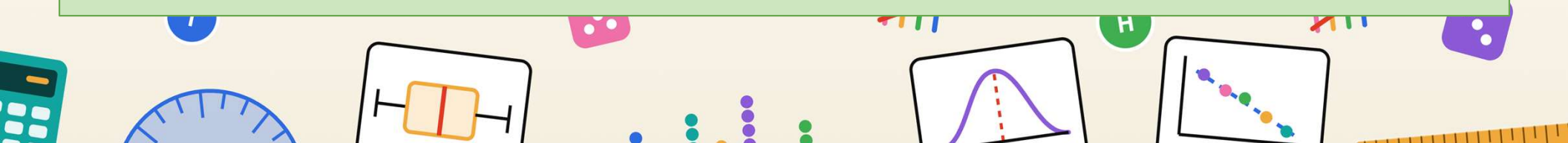
(c) Probability $\approx 0/50 = 0$, so there's about a **0% chance** it would take 23 or more boxes to get a full set. Because it is so unlikely that it would take 23 or more boxes to get a full set, this result provides convincing evidence that the 5 NASCAR drivers' cards are not equally likely to appear in each box of cereal.

Who Gets Tested?

A company randomly selects 4 employees per month for a drug test. There are **25 employees on the factory floor** and **7 in the front office**. This month, all 4 selected are from the floor. The floor workers suspect the selection wasn't fair. Estimate the probability that a fair random selection gives 4 floor workers.

(a) Describe how you would use a random number generator to carry out this simulation.

(a) Label the floor workers 1 to 25 and the office workers 26 to 32. Generate 4 DIFFERENT integers from 1 to 32 — ignore repeats, so no one gets tested twice. Record how many of the 4 are from the floor. Perform many selections; find the percentage in which all 4 are floor workers.



Who Gets Tested? The Verdict

(b) Perform 3 trials with a random number generator. Record the numbers and what they represent.

(c) In 39 of 100 trials, all 4 employees are from the factory floor. Convincing evidence the selection process was unfair?

(b) Trial 1: 2, 19, 23, 21 → 4 floor workers.
Trial 2: 27, 25, 31, 2 → 2 floor workers. Trial 3:
21, 16, 24, 32 → 3 floor workers.

(c) No — there's about a 39% chance of getting all 4 from the floor in a fair process. Because this probability isn't small (> 0.05), we do NOT have convincing evidence the selection was unfair.

AP® Exam Tip

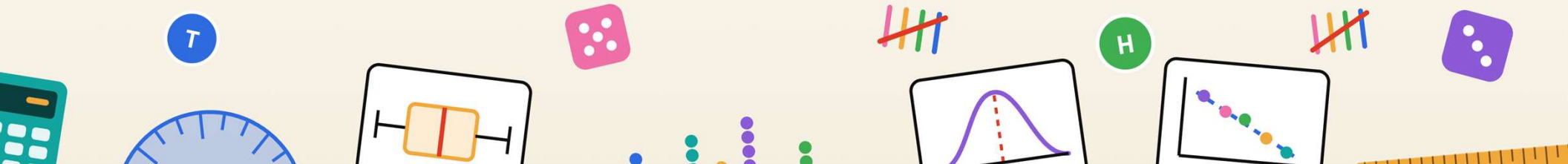
If asked to describe a simulation, your written process alone must let the reader reproduce your results. When sampling without replacement, say that repeated numbers are ignored or that different (unique) numbers are selected.

Practice: The \$15 Million Violin

Old Italian violins are supposed to sound best — some Stradivarius violins sell for more than **\$15 million**. Researchers had 10 world-class soloists play 12 violins — 6 new, 6 old (including a Stradivarius) — wearing modified welder's goggles so they couldn't tell which was which. **9 of the 10** chose a new violin as the replacement for their own instrument.

(a) Assuming the soloists are equally likely to choose any violin they tried, describe how to use a random number generator to perform one trial of a simulation.

(a) Let 1-6 = new violin and 7-12 = old violin. Generate 10 random integers from 1 to 12 to simulate the 10 soloists each choosing a violin. Record how many of the 10 integers are 1-6 (new violins).



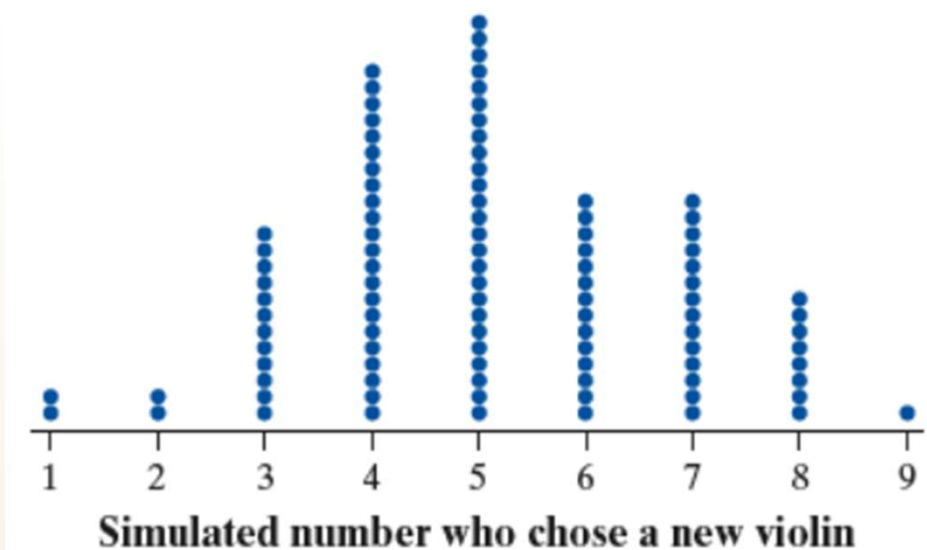
Practice: The \$15 Million Violin, Verdict

We carried out 100 trials and recorded the number of soloists (out of 10) who preferred a new violin.

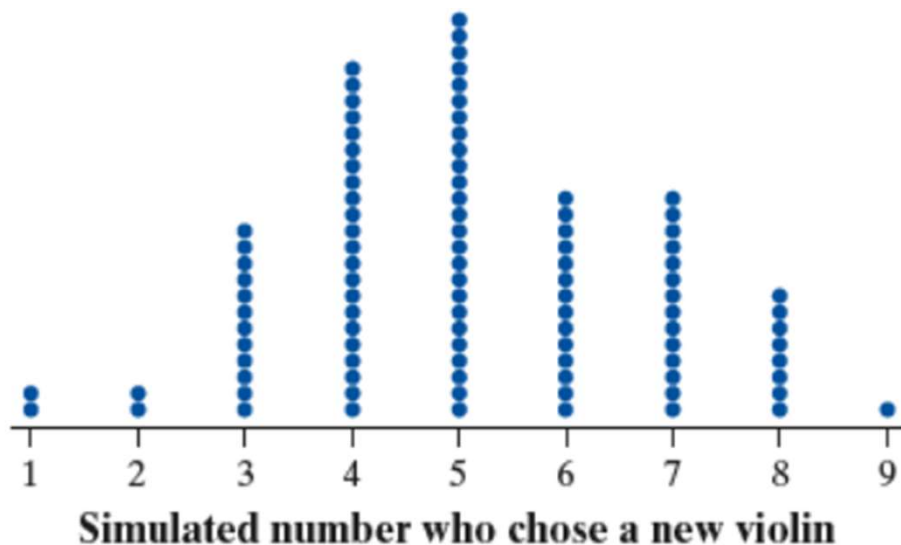
(b) Explain what the dot at 9 represents.

(c) Use the simulation results to estimate the probability that 9 or more soloists would choose a new violin.

(d) Based on the actual result (9 of 10 chose new) and your answer to (c), is there convincing evidence that new violins are more likely to be chosen?



Practice: The \$15 Million Violin, Verdict — Solutions



(b) One trial where 9 of the 10 soloists chose a new violin.

(c) 1 of the 100 dots is 9 or greater, so the probability $\approx 1/100 = 0.01$. A simulation gives an ESTIMATE — use $\approx$, not $=$.

(d) Yes. If soloists were equally likely to choose any violin, a result of 9 or more choosing new would occur only about 1% of the time by chance alone. Because $0.01 < 0.05$, there is convincing evidence that new violins are more likely to be chosen.

Two Ways to Find a Probability

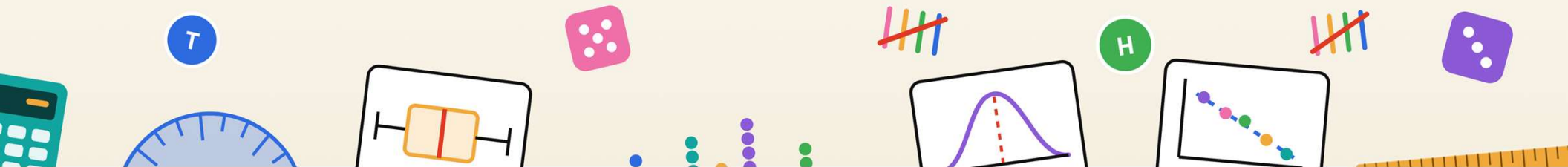
Empirical probability

Estimated from data or simulation — observe many trials and use the proportion that occurred. Flip a coin 100 times: proportion of heads ≈ 0.5 . Always an estimate: use $\approx$.

Theoretical probability

Calculated from a model of the chance process. A fair coin has two equally likely outcomes: $P(\text{heads}) = 0.5$, exactly. Use $=$.

Empirical means "learned from experience or by observation." For simple chance processes, we don't have to simulate — we can build a **model**.



Probability Models

We just used simulation to estimate probabilities. Fortunately, we don't have to always rely on simulations to determine the probability of a particular outcome.

A **probability model** is a description of some chance process that consists of two parts: a list of all possible outcomes and the probability for each outcome.



The list of all possible outcomes is called the **sample space**. A sample space can be very simple or very complex.



Practice: Identify the Sample Space

Identify the sample space for one trial of the following probability experiments:

(a) Flipping a coin

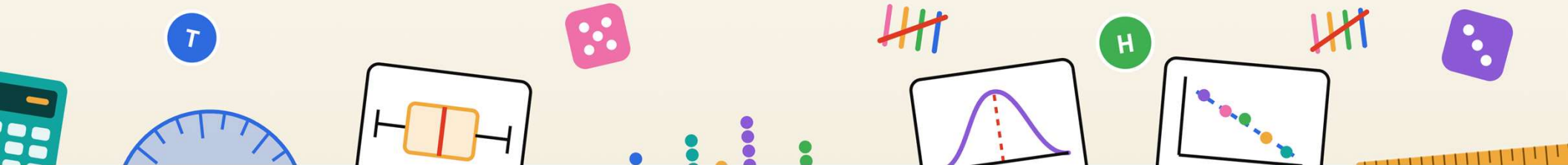
(b) Rolling a standard die

(c) Tossing 3 coins & recording the # of tails

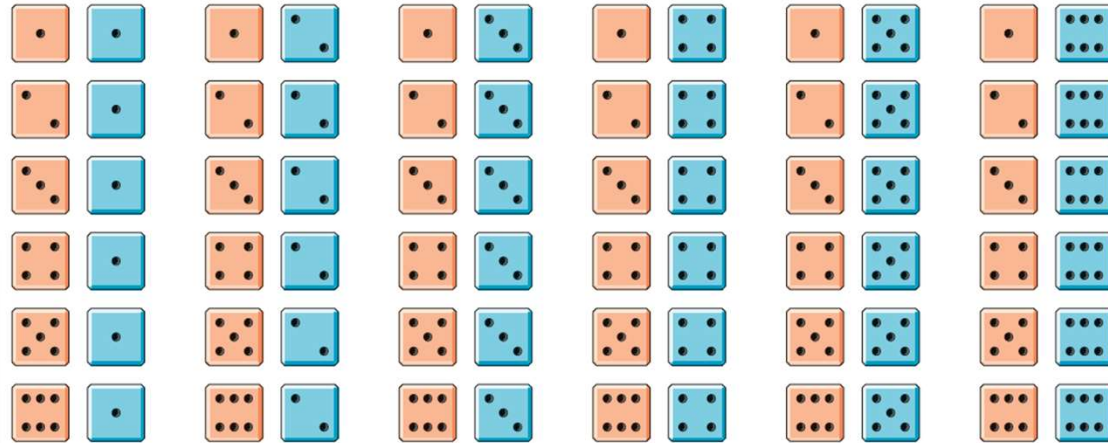
(a) $S = \{\text{Heads, Tails}\}$

(b) $S = \{1, 2, 3, 4, 5, 6\}$

(c) $S = \{0, 1, 2, 3\}$

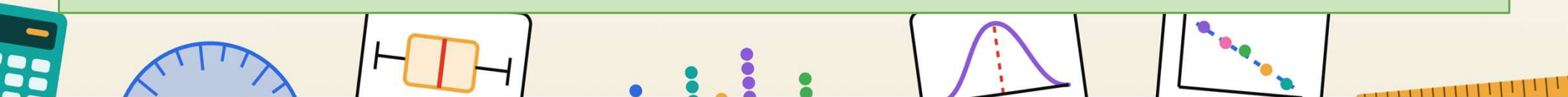


Probability Models (Rolling 2 Dice)



Sample space: 36 outcomes. Since the dice are fair, each outcome is equally likely — each outcome has probability **1/36**.

An **event** is any collection of outcomes from some chance process. We use capital letters (i.e. "A", "B", "C") to denote events.



Probability Models

(a) Let A = getting double 1's when two fair dice are rolled. Find $P(A)$, the probability of event A .

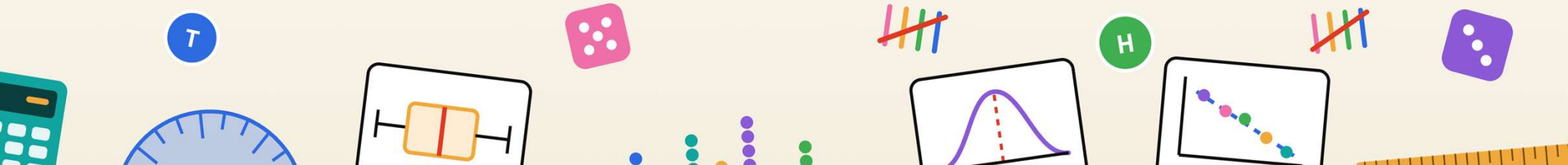
(b) Let B = rolling doubles when two fair dice are rolled. Find $P(B)$.

(c) Let C = getting a sum of 5 when two fair dice are rolled. Find $P(C)$.

(a) $P(A) = 1/36$

(b) $P(B) = 6/36 = 1/6$

(c) Hmm — how many ways can two dice sum to 5?



Probability Models

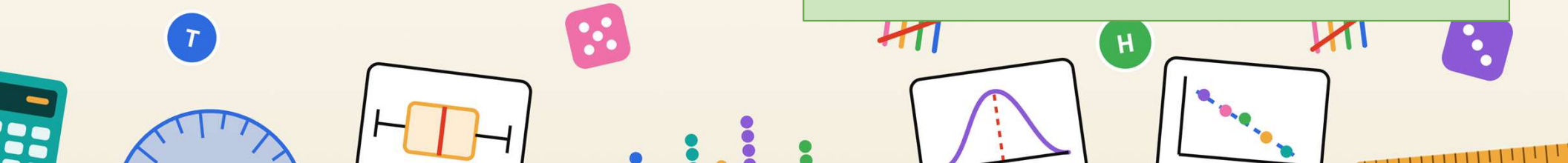
Let C = getting a sum of 5 when two fair dice are rolled. Find $P(C)$.



There are 4 outcomes that result in a sum of 5 $\rightarrow P(C) = 4/36$.

The sum of the probabilities of all the outcomes in a sample space is _____

1. $P(\text{sample space}) = 1$ — and the outcomes in a sample space are nonoverlapping.

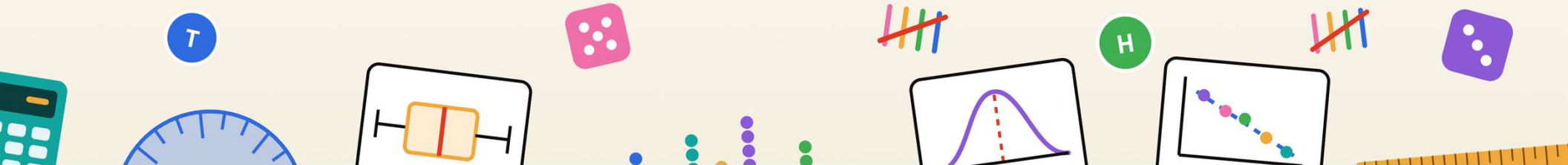


Probability Models

Finding Probabilities: Equally Likely Outcomes:

If all outcomes in the sample space are equally likely, the probability that event A occurs can be found using the formula:

$$P(A) = \frac{\text{number of outcomes in event A}}{\text{total number of outcomes in sample space}}$$



Prime Time: Build the Model Yourself

Four-sided dice are shaped like pyramids, so 1, 2, 3, and 4 are equally likely. You roll a fair four-sided die AND a fair six-sided die.

(a) Give a probability model for this random process.

(b) Let A = the sum is a **prime number** (2, 3, 5, or 7). Find $P(A)$.

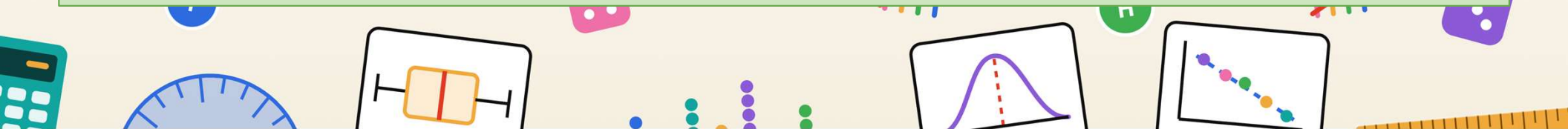
4-side / 6-side	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10

Prime Time: Build the Model Yourself — Solutions

4-side / 6-side	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10

(a) Sample space: all 24 pairs (1,1), (1,2), ... (4,6) — each with probability $1/24$. A probability model = every outcome + its probability.

(b) Prime sums: count them in the table above — 11 of the 24 pairs are 2, 3, 5, or 7. $P(A) = \mathbf{11/24}$
 $\approx \mathbf{0.458}$.



Basic Probability Rules

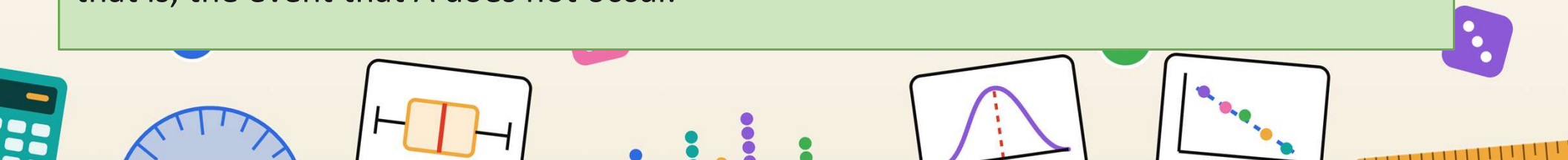
Rules that a valid probability model must obey:

- If all outcomes in the sample space are equally likely, the probability that event A occurs is

$$P(A) = \frac{\text{number of outcomes in event A}}{\text{total number of outcomes in sample space}}$$

- The probability of any event is a number between 0 and 1.
- All possible outcomes together must have probabilities that add up to 1.
- The probability that an event does *not* occur is 1 minus the probability that the event does occur.

The **complement rule** says that $P(A^c) = 1 - P(A)$, where A^c is the complement of event A; that is, the event that A does not occur.



Let's Practice

Let's work on the **Mixed Review** in your Guided Notes Packet to check your understanding of the Basic Probability Rules so far.



Basic Probability Rules

Many problems involve finding the probability of two or more events.

For example, at a large political gathering, you might wish to know, for a person selected at random, the probability the person is a female or is a Republican.

There are 3 possibilities to consider:

- 1) The person is a female
- 2) The person is a Republican
- 3) The person is both a female and a Republican

Consider another example from the same gathering. If a person is selected at random, what is the probability they are a Democrat or an Independent?

Now, there are only 2 possibilities to consider:

- 1) The person is a Democrat
- 2) The person is an Independent



Two-Way Tables, Probability, and the General Addition Rule

There are two different uses of the word 'or' in everyday life.

Soup or salad?

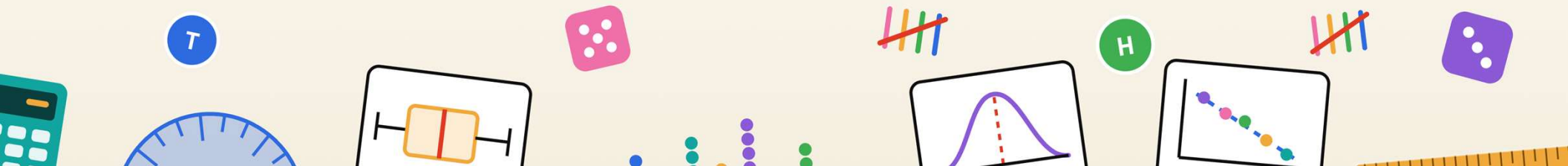
The waiter wants you to choose one or the other, but not both.

Cream or sugar?

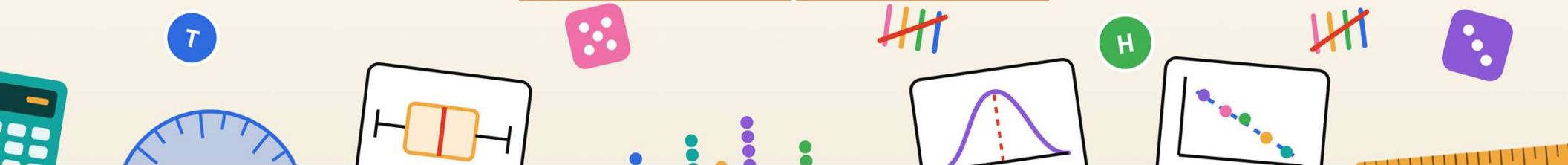
When you order coffee, it's OK to ask for one or the other — or both.

In mathematics and probability, '**A or B**' means one or the other or both.

When you're trying to find probabilities involving two events, like $P(A \cup B)$, '**A or B**', a two-way table can display the sample space in a way that makes probability calculations easier.



REMEMBER THE SPOOKY QUIZ?



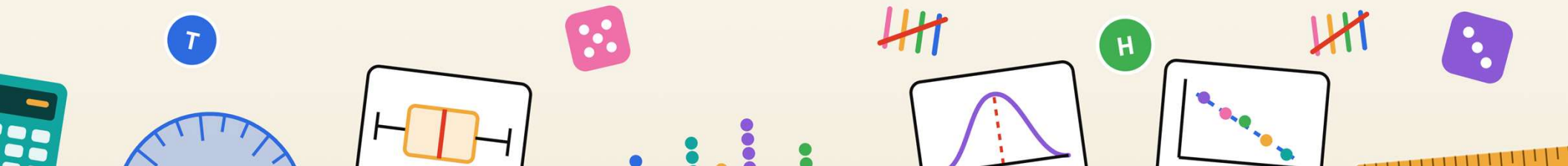
Basic Probability Rules

Two events A and B are **mutually exclusive (disjoint)** if they have no outcomes in common and so can never occur together — that is, if $P(A \cap B) = 0$ — read 'A and B.'

The addition rule for mutually exclusive (disjoint) events A and B says that $P(A \cup B) = P(A) + P(B)$.

CAUTION: this rule works ONLY for mutually exclusive (disjoint) events.

Let's practice determining if events are mutually exclusive in your Guided Notes!



Basic Probability Rules

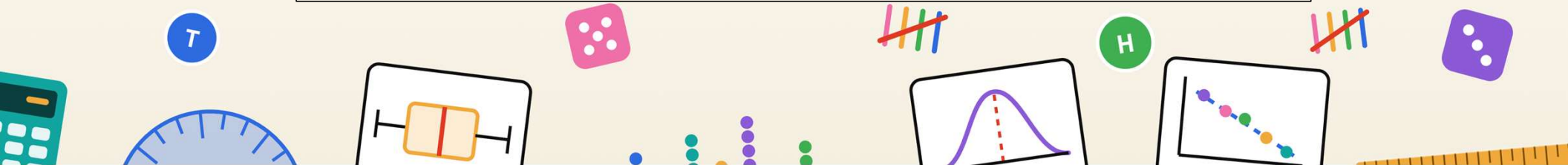
We can summarize the basic probability rules in symbolic form.

Basic Probability Rules (in symbolic form)

- For any event A , $0 \leq P(A) \leq 1$.
- If S is the sample space in a probability model, $P(S) = 1$
- In the case of equally likely outcomes,

$$P(A) = \frac{\text{number of outcomes in event } A}{\text{total number of outcomes in sample space}}$$

- **Complement rule:** $P(A^c) = 1 - P(A)$
- **Addition rule for mutually exclusive events:** If A and B are mutually exclusive, $P(A \text{ or } B) = P(A) + P(B)$



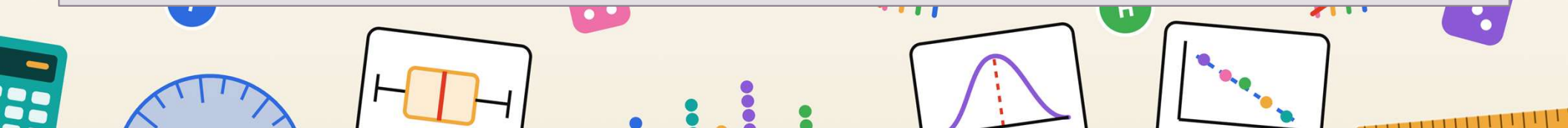
Taste the Rainbow? Wrong Candy.

Problem: Tear open a bag of M&M'S Milk Chocolate Candies, pour one candy into your hand, and observe the color. Mars, Inc. gives this probability model for its Cleveland factory:

Color	Blue	Orange	Green	Yellow	Red	Brown
Probability	0.207	0.205	0.198	0.135	0.131	0.124

- (a) Explain why this is a valid probability model.
- (b) Find the probability that you don't get a blue M&M.
- (c) What's the probability that you get an orange or a brown M&M?

$$\begin{aligned}(c) P(\text{orange or brown}) &= P(\text{orange}) + P(\text{brown}) \\ &= 0.205 + 0.124 \\ &= 0.329\end{aligned}$$



Joint Probability = Joint Relative Frequency

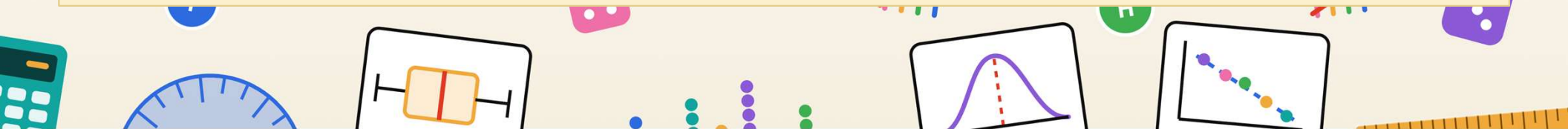
The value where a row and a column come together is a **joint probability**, $P(A \cap B)$. It's the exact same number you called a *joint relative frequency* — new name, same idea. And if $P(A \cap B) = 0$, the events are mutually exclusive.

	<\$50K	\$50–\$100K	>\$100K	Total
Very likely	184	114	84	382
Somewhat likely	84	65	41	190
Not very likely	13	27	5	45
Not at all likely	22	9	7	38
Not sure	113	55	28	196
Total	416	270	165	851

Joint Probability = Joint Relative Frequency (cont.)

	<\$50K	\$50–\$100K	>\$100K	Total
Very likely	184	114	84	382
Somewhat likely	84	65	41	190
Not very likely	13	27	5	45
Not at all likely	22	9	7	38
Not sure	113	55	28	196
Total	416	270	165	851

- $P(\text{very likely} \cap \text{income} \geq \$50\text{K}) = (114 + 84)/851 = 198/851$ — where the row meets the columns.
- Order doesn't matter: 'A and B' is the same event as 'B and A', so $P(A \cap B) = P(B \cap A)$.



Two-Way Tables, Probability, and the General Addition Rule

Problem: A college statistics class recorded gender and pierced-ear status for all 178 people in the class. Choose a student at random. Event A: getting a male student. Event B: getting a student with a pierced ear.

		Gender		
		Male	Female	Total
Pierced ear	Yes	19	84	103
	No	71	4	75
	Total	90	88	178

(a) Find $P(B)$.

(b) Find $P(A \cap B)$. Interpret this value in context.

(c) Find $P(A \cup B)$.

(c) $P(A \text{ or } B) = P(\text{male or pierced ear})$

$$= \frac{71+19+84}{178}$$

$$= 0.978$$

randomly selected student from this class is

Why does the addition rule for mutually exclusive outcomes not work for part (c)?

Two-Way Tables, Probability, and the General Addition Rule

We can fix the **double-counting** problem by subtracting $P(\text{male} \cap \text{pierced ear})$ from the sum of $P(A)$ and $P(B)$.

Outcomes here are double-counted by $P(A) + P(B)$

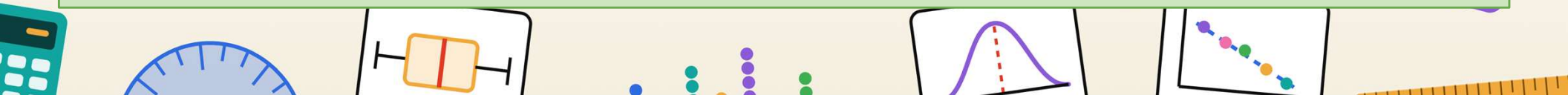
	Gender		Total	
	Male	Female		
Pierced ear	Yes	19	103	$P(B) = \frac{103}{178}$
	No	4	75	
Total	90	88	178	

$P(A \text{ and } B) = \frac{19}{178}$

$P(A) = \frac{90}{178}$

$P(A \cup B) = 90/178 + 103/178 - 19/178 = 174/178$. Lesson: we can't use the addition rule for mutually exclusive events unless A and B have no outcomes in common.

If A and B are any two events resulting from some chance process, the **general addition rule** says that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

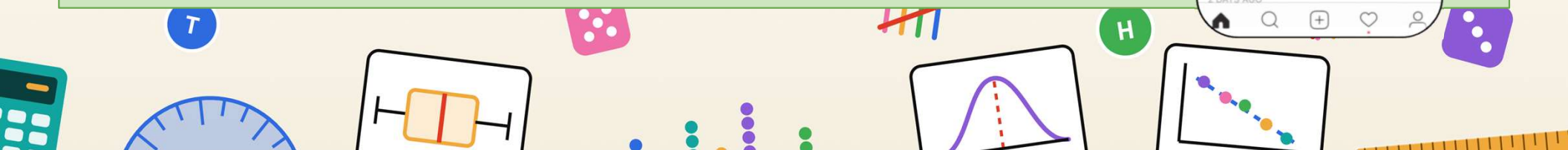


Two-Way Tables, Probability, and the General Addition Rule

A survey of all residents in a large apartment complex reveals that 68% use Facebook, 28% use Instagram, and 25% do both. Suppose we select a resident at random. What's the probability that the person uses Facebook or uses Instagram?

Let event F = uses Facebook and I = uses Instagram.

$$P(F \cup I) = P(F) + P(I) - P(F \cap I) = 0.68 + 0.28 - 0.25 = \mathbf{0.71}$$



Aw, Nuts!

During one month of production, 35% of a candy maker's chocolate bars contained nuts, 20% contained caramel, and 12% contained **both**. You grab one bar at random — what's the probability it contains nuts or caramel?

	Contain nuts	Do not contain nuts	Total
Caramel	12%	8%	20%
Not caramel	23%	57%	80%
Total	35%	65%	100%

(a) Fill in the two-way table above.

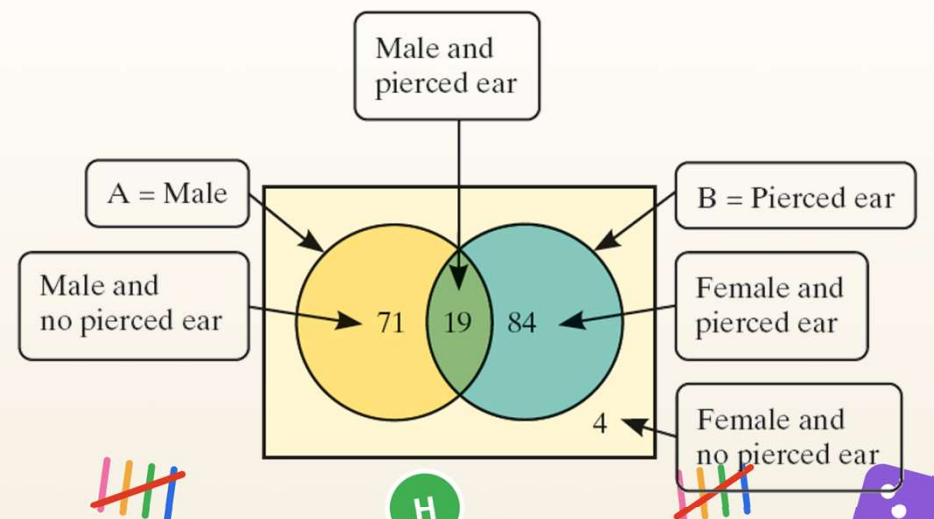
(b) Find $P(N \cup C)$.

(b) $P(N \cup C) = P(N) + P(C) - P(N \cap C) = 0.35 + 0.20 - 0.12 = 0.43$. Table check: $12\% + 23\% + 8\% = 43\%$ ✓

Venn Diagrams and Probability

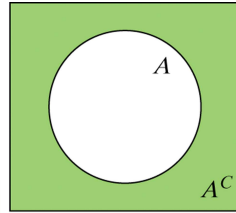
A **Venn diagram** consists of one or more circles surrounded by a rectangle. Each circle represents an event. The region inside the rectangle represents the sample space of the chance process.

Region in Venn diagram	In words	Count
In the intersection of two circles	Male and pierced ear	19
Inside circle A, outside circle B	Male and no pierced ear	71
Inside circle B, outside circle A	Female and pierced ear	84
Outside both circles	Female and no pierced ear	4

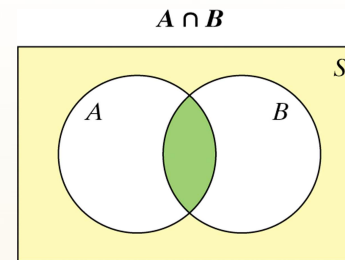


Venn Diagrams and Probability

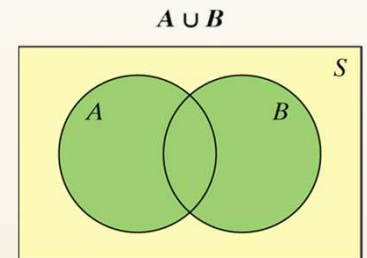
The **complement** A^C contains exactly the outcomes that are not in A .



The **intersection** of events A and B ($A \cap B$) is the set of all outcomes in both events A and B .



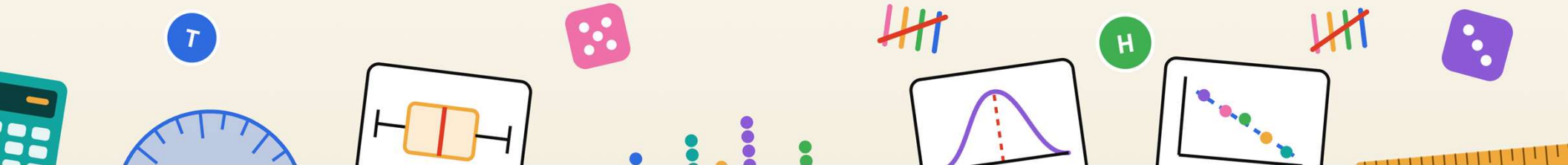
The **union** of events A and B ($A \cup B$) is the set of all outcomes in either event A or B .



HINT: To keep the symbols straight, remember

$\cup$ for **U**nion and

$\cap$ for **i**ntersection.

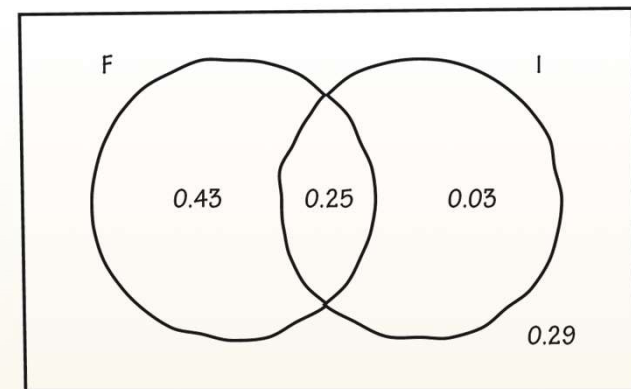


Venn Diagrams and Probability

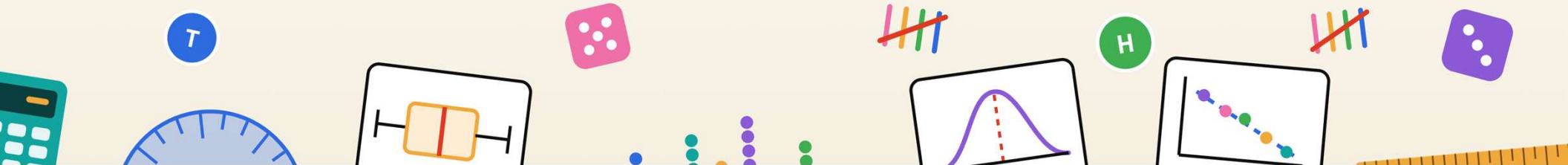
Problem: Same survey — 68% of residents use Facebook, 28% use Instagram, and 25% do both. We select a resident at random.

(a) Make a Venn diagram to display the sample space of this chance process using the events F: uses Facebook and I: uses Instagram.

(b) Find the probability that the person uses neither Facebook nor Instagram.



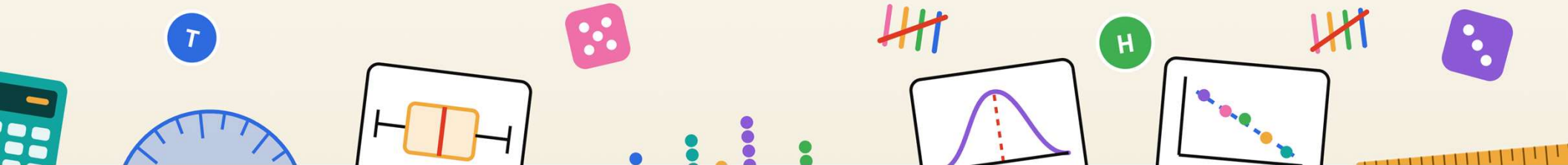
(b) $P(\text{no Facebook and no Instagram}) = P(F^C \cap I^C) = 0.29$



Understanding Check: Table or Venn — Dealer's Choice

One adult is chosen at random from the alien-technology survey. Event A: answered 'not sure.' Event B: family income under \$50,000. Use the two-way table or sketch a Venn diagram — **your pick**.

	<\$50K	\$50–\$100K	>\$100K	Total
Very likely	184	114	84	382
Somewhat likely	84	65	41	190
Not very likely	13	27	5	45
Not at all likely	22	9	7	38
Not sure	113	55	28	196
Total	416	270	165	851



Understanding Check: Table or Venn — Dealer's Choice (cont.)

(a) Find $P(A)$.

(b) Find $P(A \cap B)$.

(c) Find $P(A \cup B)$.

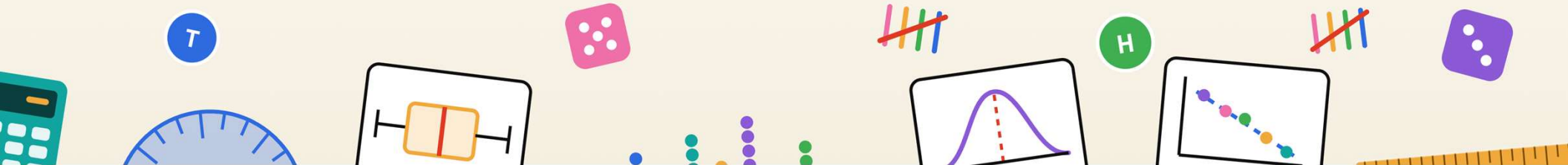
(d) Are A and B mutually exclusive? Justify using a joint probability.

(a) $P(A) = 196/851 \approx 0.230$

(b) $P(A \cap B) = 113/851 \approx 0.133$

(c) $P(A \cup B) = 196/851 + 416/851 - 113/851 = 499/851 \approx 0.586$

(d) NO. $P(A \cap B) = 113/851 \neq 0$ — they CAN happen together, so not mutually exclusive.

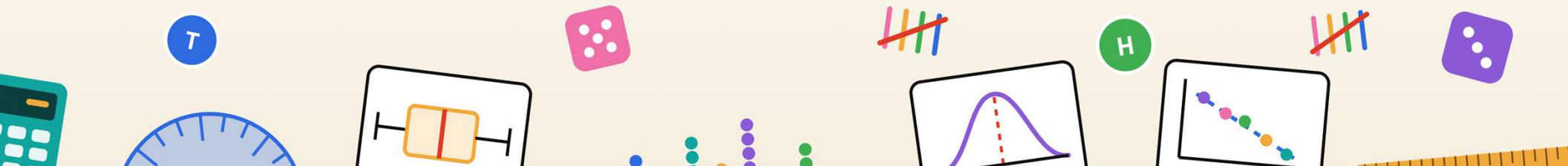


Conditional Probability

The probability we assign to an event can change if we know that some other event has occurred. This idea is the key to many applications of probability.

The probability that one event happens given that another event is known to have happened is called a **conditional probability**. The conditional probability that event B happens given that event A has happened is denoted by $P(B \mid A)$.

Read $\mid$ as “given that” or
“under the condition that”



Can You Do Both?

One student is chosen at random from a college stats class. A: can roll their tongue. B: can raise one eyebrow.

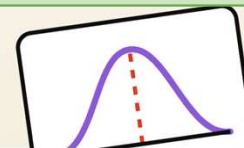
	Tongue: Yes	Tongue: No	Total
Eyeblink: Yes	42	20	62
Eyeblink: No	107	31	138
Total	149	51	200

(a) Given that the student can raise one eyebrow, what's the probability they can roll their tongue? Hint: $P(A \mid B)$

(b) Given that the student can roll their tongue, what's the probability they can raise one eyebrow? Hint: $P(B \mid A)$

$$(a) P(A \mid B) = 42/62 \approx 0.677$$

$$(b) P(B \mid A) = 42/149 \approx 0.282 \text{ -- same 42, different condition!}$$



What Is Conditional Probability?

In 1912, the luxury liner *Titanic*, on its first voyage across the Atlantic, struck an iceberg and sank. Some passengers got off the ship in lifeboats, but many died. Suppose we randomly select one of the adult passengers from the *Titanic*. Define events F: first-class passenger, S: survived, and T: third-class passenger.

		Class of travel			
		First	Second	Third	Total
Survival status	Survived	197	94	151	442
	Died	122	167	476	765
	Total	319	261	627	1207

(a) Find $P(T | S)$. Interpret this value in context.

(b) Given that the chosen person is not a first-class passenger, what's the probability that she or he survived? Write your answer as a probability statement using correct symbols for the events.

(a) $P(T|S) = P(\text{third-class} \mid \text{survived})$

(b) $P(\text{survived} \mid \text{not first-class}) = P(S|F^c)$

$$= \frac{94+1}{261+627}$$

$$= 0.276$$

Gi

chance that she or he was a third-class passenger.

What Is Conditional Probability?

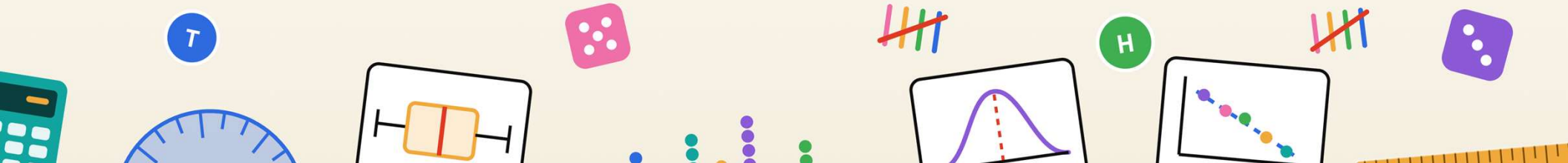
Calculating Conditional Probabilities

To find the conditional probability $P(A | B)$, use the formula

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{P(A \cap B)}{P(B)} = \frac{P(\text{both events occur together})}{P(\text{given event occurs})}$$

The choice of which event is A and which is B is arbitrary, so it is also true that:

$$P(B|A) = \frac{P(B \text{ and } A)}{P(A)} = \frac{P(B \cap A)}{P(A)}$$

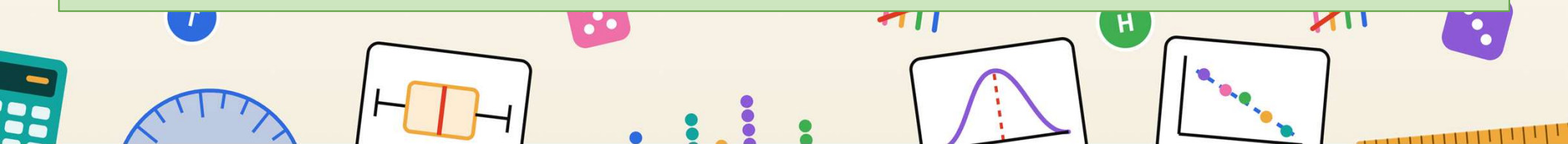


You Already Knew This

A **conditional probability** is just a *conditional relative frequency* that comes from a random process -- you have been computing these since two-way tables opened this unit.

	Tongue: Yes	Tongue: No	Total
eyebrow: Yes	0.21	0.10	0.31
eyebrow: No	0.535	0.155	0.69
Total	0.745	0.255	1.00

- Counts: $P(A \mid B) = 42/62 = 0.677$. Probabilities: $0.21/0.31 = 0.677$. Same answer.
- Joint probability divided by the condition's probability -- that's exactly the formula from the last slide.

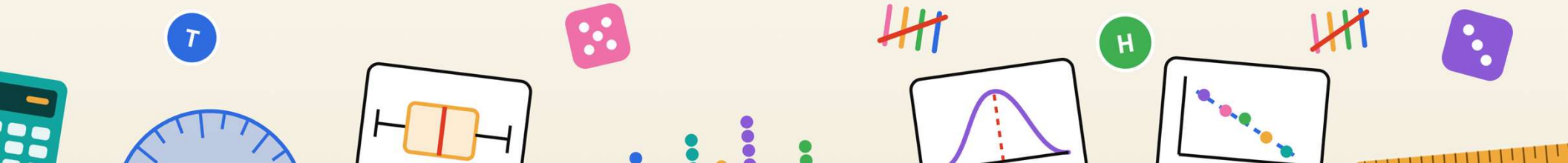


What Is Conditional Probability?

A survey of all residents in a large apartment complex reveals that 68% use Facebook, 28% use Instagram, and 25% do both. Suppose we select a resident at random.

Given that the person uses Facebook, what's the probability that she or he uses Instagram?

$$\begin{aligned} P(\text{Instagram} \mid \text{Facebook}) &= P(I \mid F) \\ &= \frac{P(I \cap F)}{P(F)} \\ &= \frac{0.25}{0.68} \\ &= 0.368 \end{aligned}$$



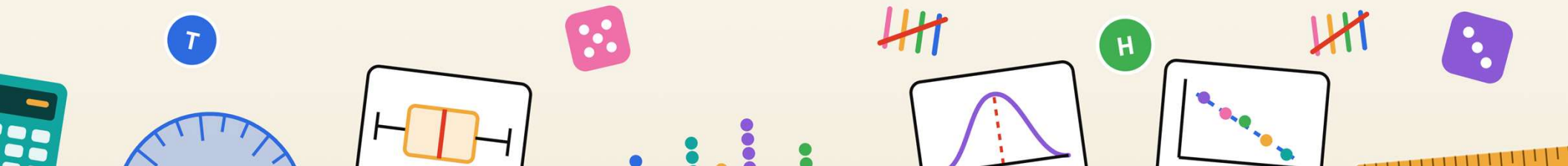
No Memory: Independent Events

Toss a fair coin twice. A: first toss is a head. B: second toss is a head. The coin has no memory: $P(B \mid A) = 1/2$ and $P(B \mid A^c) = 1/2$. Knowing how the first toss landed tells you nothing about the second.

$$P(B \mid A) = P(B \mid A^c) = P(B) = 1/2 \rightarrow A \text{ and } B \text{ are independent events.}$$

AP® Exam Tip

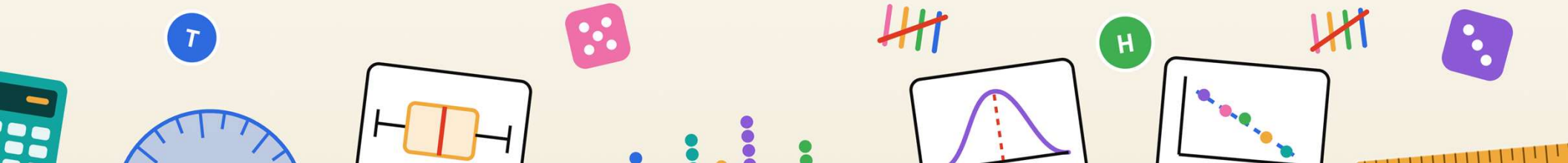
The formulas for determining whether two events are independent are NOT on the AP Statistics formula sheet -- know them cold.



Conditional Probability and Independence

A and B are **independent events** if knowing whether or not one event has occurred does not change the probability that the other event will happen. In other words, events A and B are independent if $P(A \mid B) = P(A \mid B^c) = P(A)$. Alternatively, events A and B are independent if $P(B \mid A) = P(B \mid A^c) = P(B)$.

		Gender		
		Male	Female	Total
Pierced ear	Yes	19	84	103
	No	71	4	75
	Total	90	88	178



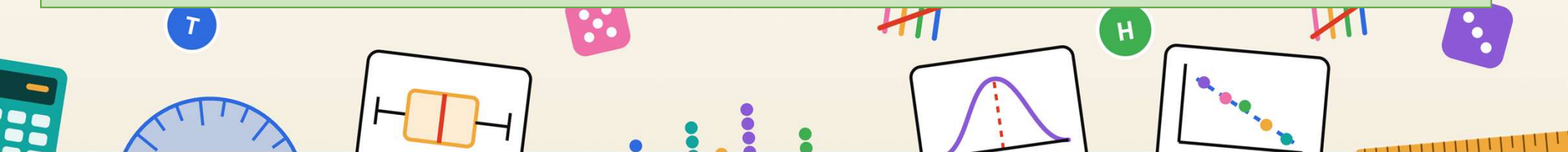
Conditional Probability and Independence — Solutions

		Gender		
		Male	Female	Total
Pierced ear	Yes	19	84	103
	No	71	4	75
	Total	90	88	178

$$P(\text{pierced ear} \mid \text{male}) = P(B \mid A) = 19/90 = 0.211$$

$$P(\text{pierced ear} \mid \text{female}) = P(B \mid A^c) = 84/88 = 0.954$$

Knowing that the chosen student is a male changes (greatly reduces) the probability that the student has a pierced ear. So these two events are **not independent**.



Conditional Probability and Independence

Is there a relationship between gender and handedness? To find out, we used Census At School's Random Data Selector to choose an SRS of 100 Australian high school students who completed a survey. The two-way table summarizes gender and dominant hand. Suppose we choose one of the students in the sample at random.

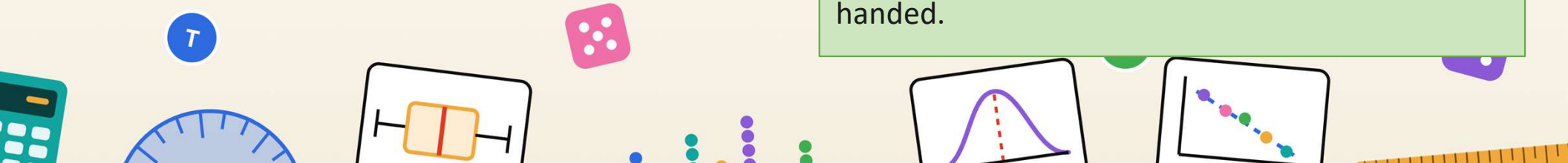
		Gender		
		Male	Female	Total
Dominant hand	Right	39	51	90
	Left	7	3	10
	Total	46	54	100

Are the events 'male' and 'left-handed' independent? Justify your answer.

$$P(\text{left-handed} \mid \text{male}) = 7/46 = 0.152$$

$$P(\text{left-handed} \mid \text{female}) = 3/54 = 0.056$$

Because these probabilities are not equal, the events 'male' and 'left-handed' are **not independent**. Knowing that the student is male increases the probability that the student is left-handed.

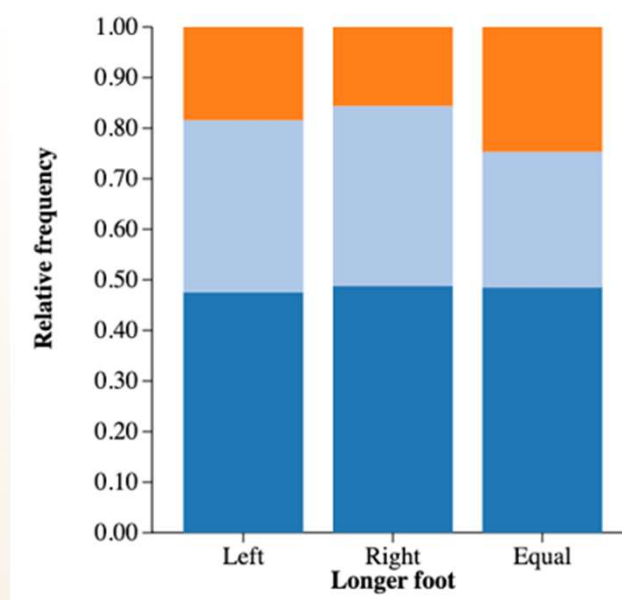


Independence Means No Association

Longer foot vs. longer finger for a random sample of 400 seniors. If events from a two-way table are **independent**, the variables have **no association** -- every bar in the segmented bar graph splits the same way.

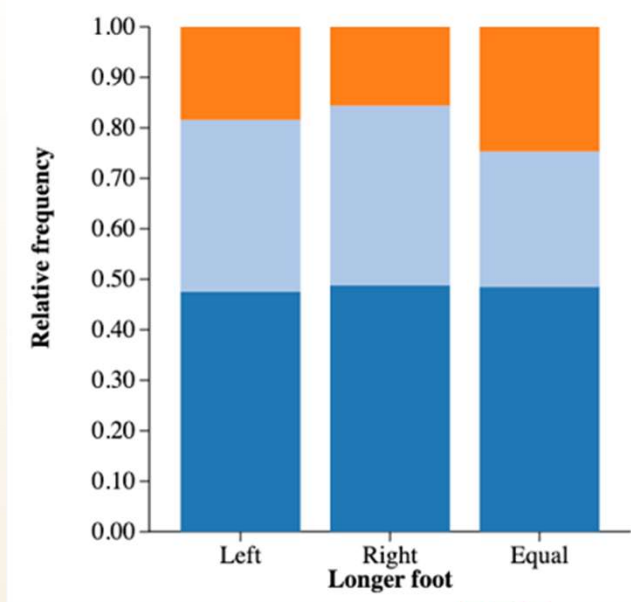
Is there an association between longer foot and longer finger in this *sample*?

The segments differ slightly from bar to bar → a slight association → not quite independent in the sample.



Independence Means No Association

Longer foot vs. longer finger for a random sample of 400 seniors. If events from a two-way table are **independent**, the variables have **no association** -- every bar in the segmented bar graph splits the same way.



Does that prove an association in the *population* of all seniors who took the survey?

The segments differ slightly from bar to bar → a slight association → not quite independent in the sample.

The General Multiplication Rule

About 55% of high school students play on a school athletic team, and roughly 6% of these athletes go on to play in the NCAA.

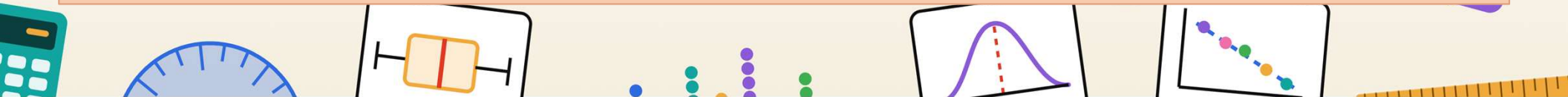
What percent of high school students play a sport in high school and go on to play on an NCAA team?

$$\begin{aligned} P(\text{high school sport and NCAA team}) &= P(\text{high school sport}) \cdot P(\text{NCAA team} \mid \text{high school sport}) \\ &= (0.55)(0.06) = 0.033 \end{aligned}$$

For any chance process, the probability that events A and B both occur can be found using the **general multiplication rule**: $P(A \cap B)$ -- 'A and B' = $P(A) \cdot P(B \mid A)$.

AP® Exam Tip

The general multiplication rule is NOT on the AP Statistics formula sheet -- know it cold.

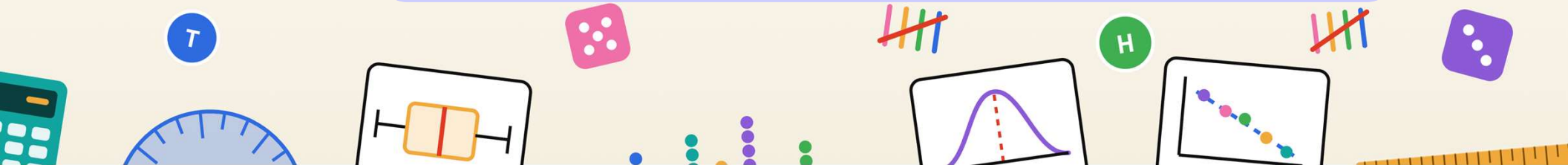


The General Multiplication Rule

The Pew Internet and American Life Project reported that 79% of teenagers (ages 13 to 17) use social media, and that 39% of teens who use social media feel pressure to post content that will be popular and get lots of comments or likes.

Find the probability that a randomly selected teen uses social media and feels pressure to post content that will be popular and get lots of comments or likes.

$$\begin{aligned} &P(\text{use social media and feel pressure}) \\ &= P(\text{use social media}) \cdot P(\text{feel pressure} \mid \text{use social media}) \\ &= (0.79)(0.39) \\ &= 0.308 \end{aligned}$$



Candy Grab: No Put-Backs

A bag holds 20 candies: 15 cherry, 5 lemon. You grab 3 at random, eating as you go -- no putting candy back.

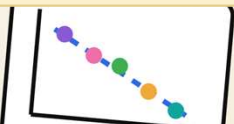
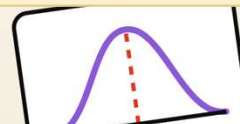
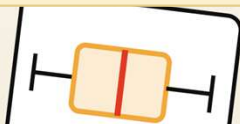
(a) $P(\text{all three cherry})$?

(b) Practice: $P(\text{no cherry at all})$?

(a) $P(\text{all cherry}) = (15/20)(14/19)(13/18) \approx 0.399$ -- the condition updates after every grab.

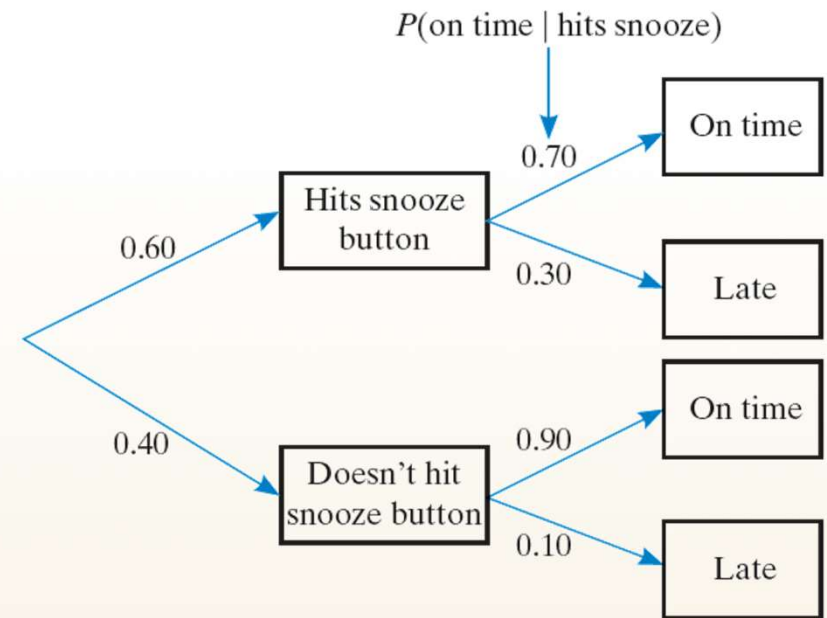
(b) $P(\text{no cherry}) = (5/20)(4/19)(3/18) \approx 0.009$ - a lemon wipeout is rare.

- Think in stages: each draw's probability is conditional on what's already gone.
- Always ask: *with* or *without* replacement? That one question drives every sampling problem to come.



Tree Diagrams and Conditional Probability

A **tree diagram** shows the sample space of a chance process involving multiple stages. The probability of each outcome is shown on the corresponding branch of the tree. All probabilities after the first stage are conditional probabilities.

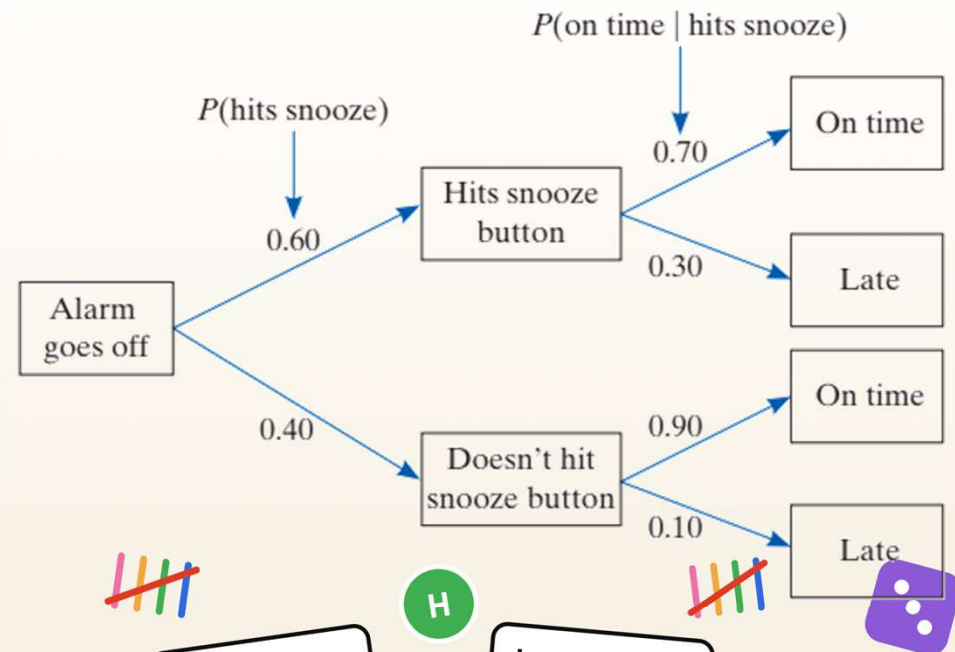


Snooze You Lose

Shannon hits snooze on 60% of school days. If she snoozes, she makes it to first period on time with probability 0.70. If she gets right up: 0.90. Select a school day at random.

Find $P(\text{snooze} \cap \text{late})$ -- 'snooze and late'.

$P(\text{snooze} \cap \text{late}) = (0.60)(0.30) = 0.18$.
Multiply along the branches.

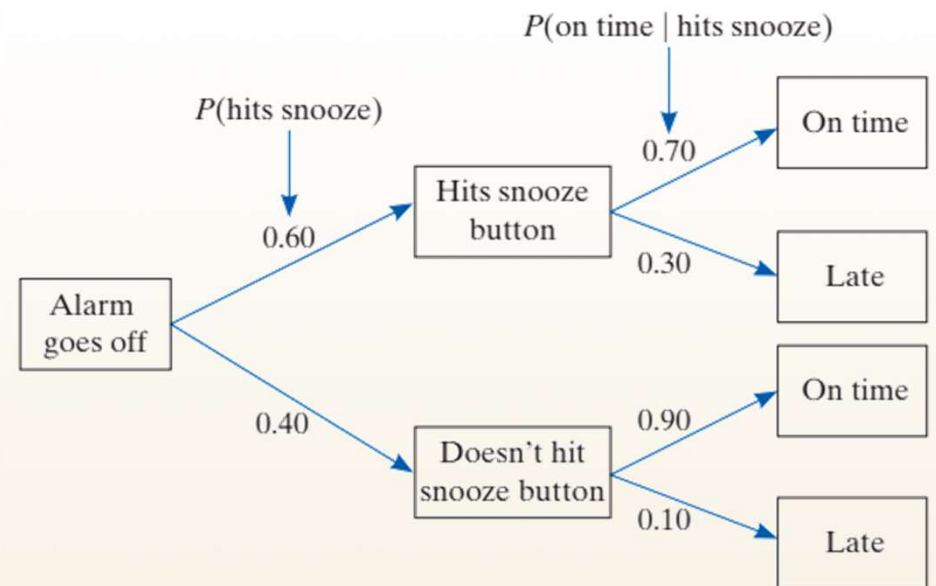


How Often Is Shannon Late?

What's $P(\text{late})$? Shannon can be late two ways -- snooze-and-late or no-snooze-and-late -- and those branches are mutually exclusive.

$$P(\text{no snooze} \cap \text{late}) = (0.40)(0.10) = 0.04$$

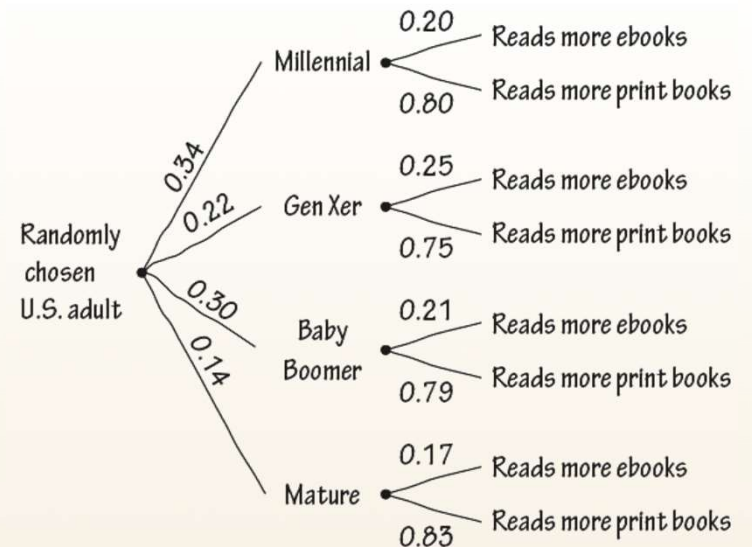
$$P(\text{late}) = 0.18 + 0.04 = 0.22 \text{ -- multiply along branches, add across them.}$$



Tree Diagrams and Conditional Probability

Harris Interactive reported that 20% of millennials, 25% of Gen Xers, 21% of baby boomers, and 17% of matures (age 68 and older) read more ebooks than print books. Per the U.S. Census Bureau, 34% of those 18 and over are millennials, 22% are Gen Xers, 30% are baby boomers, and 14% are matures. We select one U.S. adult at random and record generation and whether she or he reads more ebooks or print books.

(a) Draw a tree diagram to model this chance process.



Tree Diagrams and Conditional Probability

(b) Find the probability that the person reads more ebooks than print books.



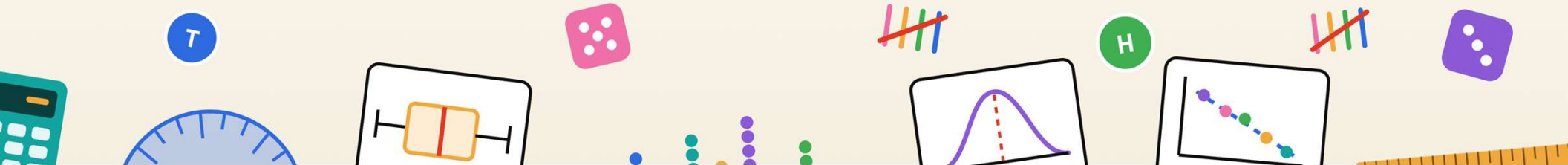
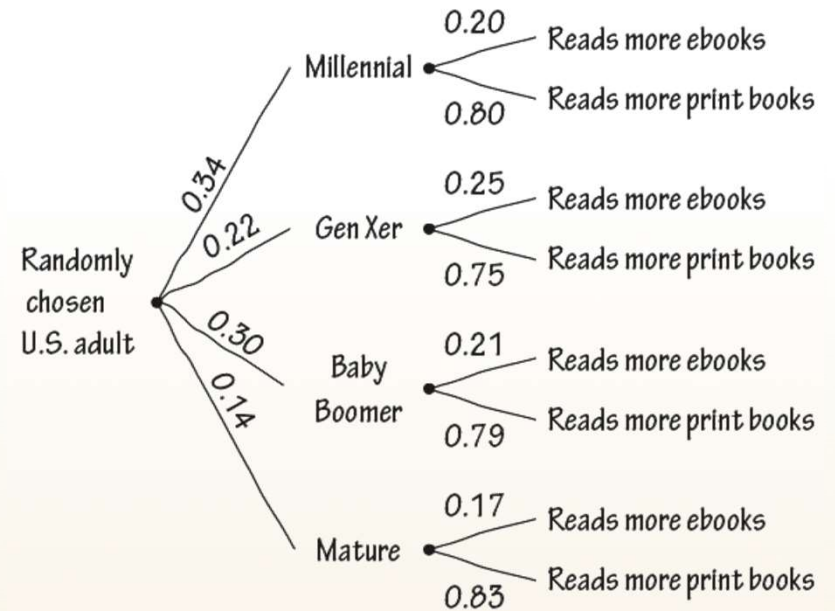
(b) $P(\text{reads more ebooks})$

$$\begin{aligned} &= (0.34)(0.20) + (0.22)(0.25) + \\ &\quad (0.30)(0.21) + (0.14)(0.17) \\ &= 0.0680 + 0.0550 \\ &\quad + 0.0630 + 0.0238 \\ &= 0.2098 \end{aligned}$$

Tree Diagrams and Conditional Probability

(c) Suppose the chosen person reads more ebooks than print books. What's the probability that she or he is a millennial?

(c) $P(\text{millennial} \mid \text{reads more ebooks}) = 0.0680 / 0.2098 \approx \mathbf{0.324}$

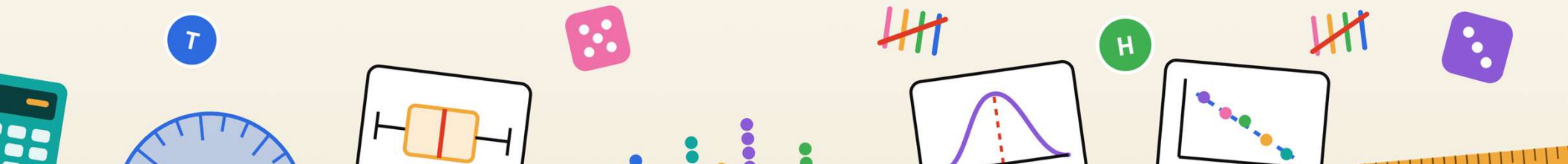
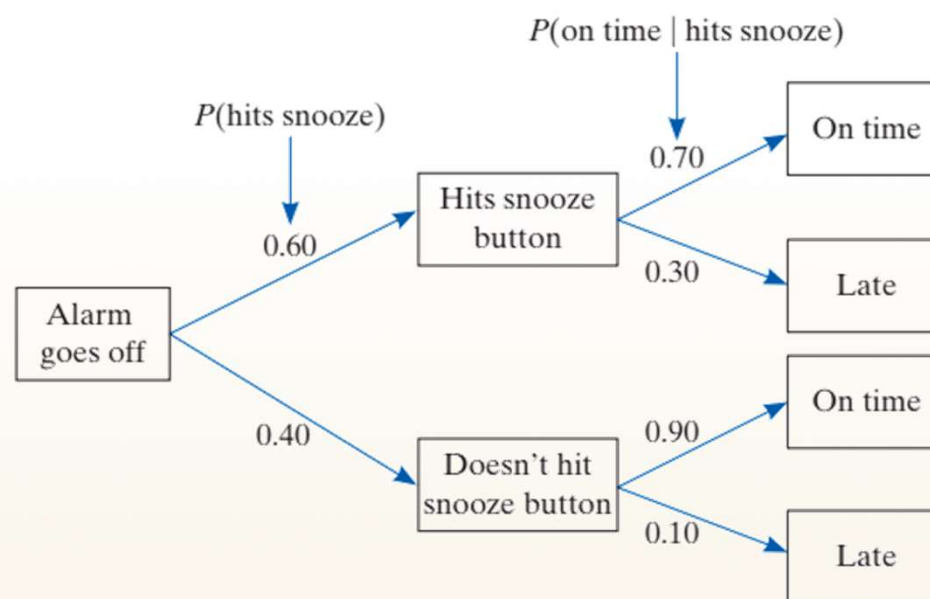


Going in Reverse

Shannon was late today. What's the probability she hit snooze? The condition now sits at the *end* of the tree -- we're driving it backward.

$$\begin{aligned} P(\text{snooze} \mid \text{late}) \\ &= P(\text{snooze} \cap \text{late}) / P(\text{late}) \\ &= 0.18 / 0.22 \approx 0.818 \end{aligned}$$

Same conditional probability formula -- the tree hands you both pieces.



Covid Scent Dogs

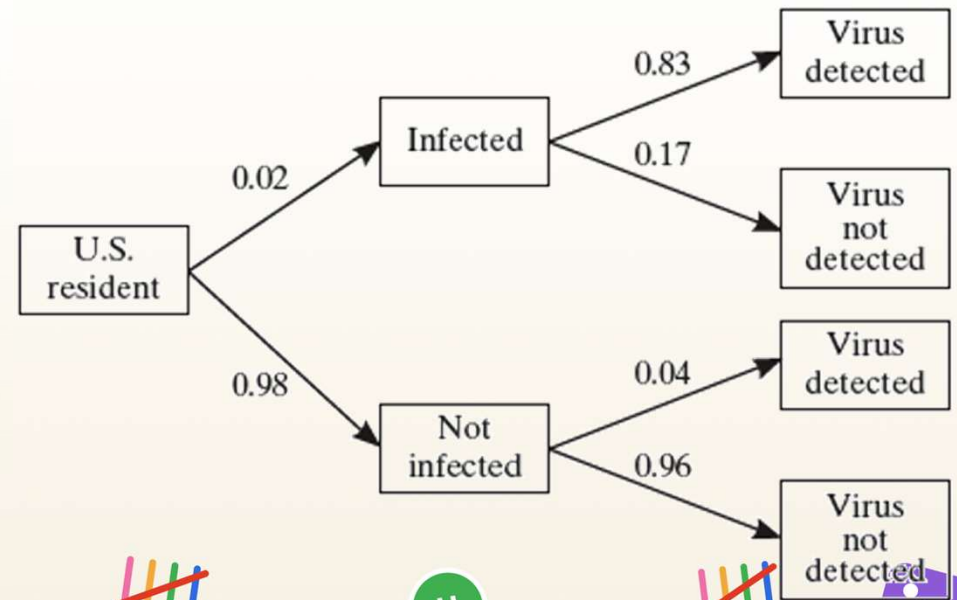
Before any math: the dog says 'infected.' Gut check -- how sure are you that this person really has the virus?

July 2020: dogs were trained to detect Covid-19 in human saliva. They alerted on 83% of infected subjects (true positive) and stayed quiet on 96% of uninfected subjects (true negative). About 2% of U.S. residents were infected. A random resident gets sniffed -- and the dog alerts.

Find $P(\text{infected} \mid \text{virus detected})$

$$P(\text{detected}) = (0.02)(0.83) + (0.98)(0.04) \\ = 0.0166 + 0.0392 = 0.0558$$

$$P(\text{infected} \mid \text{detected}) = 0.0166 / 0.0558 \approx 0.297$$



The Multiplication Rule for Independent Events

When events A and B are independent, $P(B \mid A) = P(B)$. We can simplify the general multiplication rule as follows:

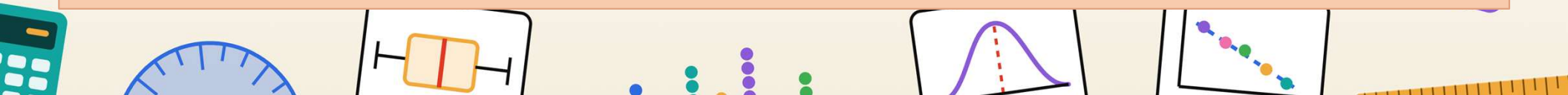
$$P(A \cap B) = P(A) \cdot P(B \mid A) = P(A) \cdot P(B)$$

If A and B are **independent events**, the probability that A and B both occur is $P(A \cap B)$ -- 'A and B' = $P(A) \cdot P(B)$.

- CAUTION: Note that this rule applies *only* to independent events.

AP® Exam Tip

The multiplication rule for independent events is NOT on the AP Statistics formula sheet -- know it cold.

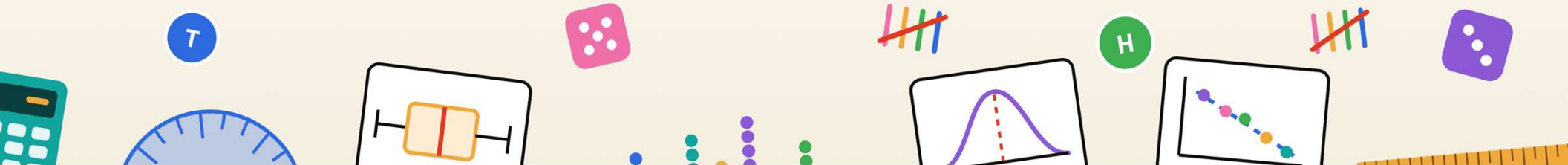


Triple Defect

A factory produces two types of products: Type A and Type B. The probability that a randomly selected Type A product is defective is 0.1, and the probability that a randomly selected Type B product is defective is 0.05.

You select three Type A products independently. What is the probability that all three products are defective?

Independent selections $\rightarrow$ multiply: $P(\text{all three defective}) = (0.1)(0.1)(0.1) = \mathbf{0.001}$



The Multiplication Rule for Independent Events

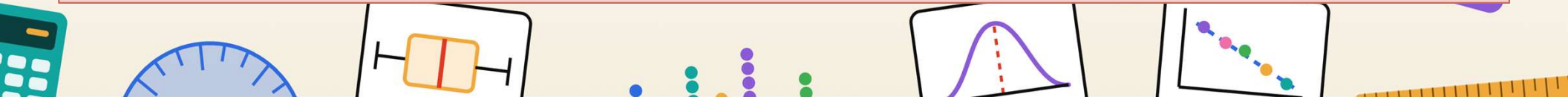
Many people who visit clinics to be tested for HIV don't come back to learn their test results, so clinics now use 'rapid HIV tests' that give a result while the client waits. The trade-off for fast results: rapid tests are less accurate than slower laboratory tests. Applied to people who do not have HIV, one rapid test has a probability of about 0.004 of producing a false positive.

If a clinic tests 200 randomly selected people who are free of HIV antibodies, what is the probability that at least one false positive will occur? Assume that test results for different individuals are independent.

$$P(\text{no false positives}) = P(\text{all 200 tests negative}) \\ = (0.996)(0.996) \cdots (0.996) = 0.996^{200} = 0.4486$$

$$P(\text{at least one false positive}) = 1 - 0.4486 = \\ \mathbf{0.5514}$$

- Pro tip! When asked for $P(\text{at least one})$, $1 - P(\text{none})$ saves a lot of calculation.



Mutually Exclusive $\neq$ Independent

Choose a U.S. resident at random. A: the person is younger than 10. B: the person has a driver's license.

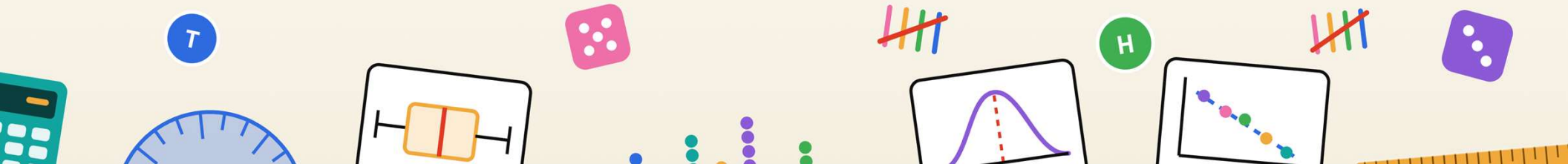
Mutually exclusive?

Yes -- they can't happen together, so $P(A \cap B) = 0$.

Independent?

No! If A happens, B is impossible: $P(B \mid A) = 0 \neq P(B)$.

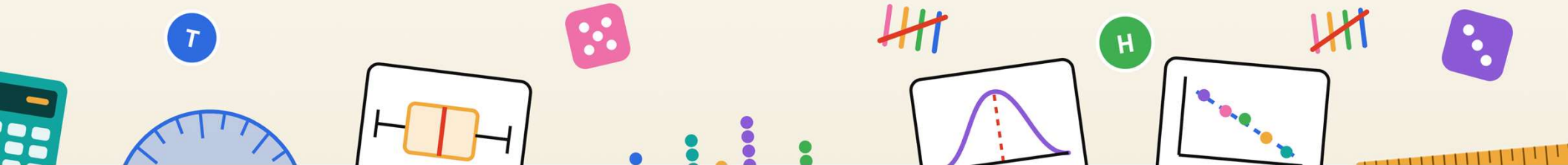
- Mutually exclusive events with nonzero probabilities can **never** be independent -- one happening *guarantees* the other doesn't.



The Tragic Case of Sally Clark

England, 1999: Sally Clark was convicted of murdering her two infant sons after an expert testified that two SIDS deaths in one family had probability $(1/8543)^2$ -- about 1 in 73 million.

- Squaring $1/8543$ treats the two deaths as **independent** events.
- But SIDS risk runs in families -- shared genetics and environment make a second tragedy far more likely than the square suggests.
- Her conviction was overturned in 2003 -- after more than three years in prison.



Understanding Check: Sports and Jobs

200 students at a school. One is chosen at random. S: plays a school sport. J: has a part-time job.

- (a)** Find $P(J \mid S)$.
- (b)** Are S and J independent?
- (c)** Two different students are chosen at random. Find $P(\text{both play a sport})$.
- (d)** Find $P(S \cup J)$.

	Job: Yes	Job: No	Total
Sport: Yes	30	50	80
Sport: No	60	60	120
Total	90	110	200

(a) $P(J \mid S) = 30/80 = 0.375$

(b) $P(J) = 90/200 = 0.45 \neq 0.375$
 $\rightarrow$ not independent.

(c) Without replacement:
 $(80/200)(79/199) \approx 0.159$

(d) $P(S \cup J) = (80 + 90 - 30)/200$
 $= 140/200 = 0.70$

