Introduction

Data can be presented in many different ways. In the previous lesson, you worked with data in conditional probabilities to interpret the relationships between various events. When you are presented with a great deal of data, two-way frequency tables can provide a convenient way to compare the frequency of data items and see the relationships among them.

Key Concepts

- Previously, you learned that there are two equivalent expressions for the conditional probability of *B* given *A*:

$$P(B|A) = \frac{P(A \text{ and } B)}{P(A)} = \frac{\text{number of outcomes in } (A \text{ and } B)}{\text{number of outcomes in } A}$$

- Conditional probability can be used to test for independence.
- Events *A* and *B* are independent events if $P(B|A) = P(B)$ or if $P(A|B) = P(A)$. (Note that if one is true, then the other is also true.)

Key Concepts, *continued*

- Remember that for real-world data, modified tests for independence are sometimes used:
  - Events *A* and *B* are independent if the occurrence of *A* has no *significant* effect on the probability of *B*; that is, $P(B|A) \approx P(B)$.
  - Events *A* and *B* are independent if the occurrence of *B* has no *significant* effect on the probability of *A*; that is, $P(A|B) \approx P(A)$.

Key Concepts, *continued*

- If A and B are two events from a sample space with $P(A) \neq 0$, then the conditional probability of B given A in set notation is $P(B|A) = \dfrac{P(A \cap B)}{P(A)}$.
- The conditional probability formula can be solved to obtain this formula for $P(A \text{ and } B)$:

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

Key Concepts, *continued*

- A **two-way frequency table** is a frequency table that shows two categories of characteristics, one in rows and the other in columns. Each cell value is a frequency that shows how many times two different characteristics appear together, or how often characteristics are associated with a person, object, or type of item that is being studied.

Key Concepts, *continued*

- The example below shows a typical setup of a two-way frequency table.

Category 1 of characteristics	Category 2 of characteristics	
	Characteristic 1	Characteristic 2
Characteristic 1	*a*	*b*
Characteristic 2	*c*	*d*

Key Concepts, *continued*

- When probabilities and conditional probabilities are calculated from data in a two-way frequency table, then totals of characteristics are needed.

Category 1 of characteristics	Category 2 of characteristics		Total
	Characteristic 1	Characteristic 2	
Characteristic 1	*a*	*b*	
Characteristic 2	*c*	*d*	
Total			

Key Concepts, *continued*

- When a two-way frequency table is used to find probabilities and conditional probabilities, the characteristics represent events, and the frequencies are numbers of outcomes.

Common Errors/Misconceptions

- using a row total in a two-way frequency table when a column total should be used, and vice versa
- using a row or column total when the sample space total should be used

Guided Practice

Example 1

The Student Council wants to host a school-wide activity. Council members survey 40 students, asking them to choose either a field trip, a dance, or a talent show. The table on the following slides shows the survey results, with the surveyed students numbered 1–40. Construct a two-way frequency table to summarize the data.

Guided Practice: Example 1, *continued*

Student	Grade	Activity
1	10	FT
2	12	D
3	10	TS
4	10	FT
5	11	D
6	12	D
7	10	TS
8	10	FT
9	10	FT
10	11	TS

Student	Grade	Activity
11	12	D
12	10	TS
13	11	TS
14	10	FT
15	11	D
16	10	FT
17	12	D
18	10	FT
19	12	D
20	11	TS

Key: TS = Talent show, FT = Field trip, D = Dance

(*continued*)

Guided Practice: Example 1, *continued*

Student	Grade	Activity
21	10	D
22	10	FT
23	12	D
24	11	D
25	11	TS
26	12	D
27	12	D
28	10	D
29	11	D
30	11	D

Student	Grade	Activity
31	12	FT
32	10	TS
33	12	D
34	11	D
35	11	FT
36	11	FT
37	11	TS
38	12	TS
39	11	FT
40	12	TS

Key: TS = Talent show, FT = Field trip, D = Dance

Guided Practice: Example 1, *continued*

1. Set up a tally table.

There are two characteristics associated with each student: that student's grade and that student's choice of activity. Set up a table that shows "Grade" and "Activity choice" as categories, and all the different characteristics in each category.

Guided Practice: Example 1, *continued*

Grade	Activity choice		
	Talent show	Field trip	Dance
10			
11			
12			

Guided Practice: Example 1, *continued*

2. Tally the data.

For each student, draw a tally mark that corresponds to that student's grade and choice of activity in the appropriate cell of the data table. The tally marks for students 1–5 are shown in this incomplete tally table.

Grade	Activity choice		
	Talent show	Field trip	Dance
10	|	||	
11			|
12			|

Guided Practice: Example 1, *continued*

The complete tally table below shows the tally marks for all the students.

Grade	Activity choice		
	Talent show	Field trip	Dance
10	IIII	~~IIII~~ III	II
11	~~IIII~~	III	~~IIII~~ I
12	II	I	~~IIII~~ IIII

Guided Practice: Example 1, *continued*

3. Create a two-way frequency table.

Count the tally marks in each cell of your tally table. Then, create another table (a two-way frequency table) to show your count results. These results are frequencies. The completed two-way frequency table is shown on the next slide.

Guided Practice: Example 1, *continued*

Grade	Activity choice		
	Talent show	**Field trip**	**Dance**
10	4	8	2
11	5	3	6
12	2	1	9

Add all the frequencies; verify that their sum is 40 (since 40 students were surveyed).

$$4 + 8 + 2 + 5 + 3 + 6 + 2 + 1 + 9 = 40$$

✔

Guided Practice: Example 1, *continued*

Guided Practice

Example 2

The completed two-way frequency table from Example 1 is shown below. It shows the results of a survey designed to help the Student Council choose a school-wide activity.

Grade	Activity choice		
	Talent show	**Field trip**	**Dance**
10	4	8	2
11	5	3	6
12	2	1	9

Guided Practice: Example 2, *continued*

Consider the following events that apply to a random student who participated in the survey.

TEN: The student is in the tenth grade.

TWELVE: The student is in the twelfth grade.

FT: The student prefers a field trip.

TS: The student prefers a talent show.

Guided Practice: Example 2, *continued*

Compare $P(TEN|FT)$ and $P(FT|TEN)$. Are *TEN* and *FT* independent?

Compare $P(TWELVE|TS)$ and $P(TS|TWELVE)$. Are *TWELVE* and *TS* independent?

Interpret the results.

Guided Practice: Example 2, *continued*

1. Find the totals of all the categories.

Grade	Activity choice			Total
	Talent show	Field trip	Dance	
10	4	8	2	14
11	5	3	6	14
12	2	1	9	12
Total	11	12	17	40

Guided Practice: Example 2, *continued*

2. Compare $P(TEN|FT)$ and $P(FT|TEN)$.

$$P(TEN|FT)=\frac{8}{12}\approx 0.667$$

There were 12 votes for a field trip; 8 were by tenth graders.

$$P(FT|TEN)=\frac{8}{14}\approx 0.571$$

There were 14 votes by tenth graders; 8 were for a field trip.

0.667 > 0.571; therefore, $P(TEN|FT)>P(FT|TEN)$.

Guided Practice: Example 2, *continued*

3. Determine if *TEN* and *FT* are independent.

Remember that events *A* and *B* are independent events if $P(B|A) = P(B)$ or if $P(A|B) = P(A)$.

Compare $P(TEN|FT)$ with $P(TEN)$ and $P(FT|TEN)$ with $P(FT)$.

Guided Practice: Example 2, *continued*

$$P(TEN|FT)=\frac{8}{12}\approx 0.667$$

There were 12 votes for a field trip; 8 were by tenth graders.

$$P(TEN)=\frac{14}{20}=0.35$$

There were 40 votes in all; 14 were by tenth graders.

$0.667 \neq 0.35$; therefore, $P(TEN|FT) \neq P(TEN)$.

Guided Practice: Example 2, *continued*

$$P(FT|TEN) = \frac{8}{14} \approx 0.571$$

There were 14 votes by tenth graders; 8 were for a field trip.

$$P(FT) = \frac{12}{40} = 0.3$$

There were 40 votes in all; 12 were for a field trip.

$0.571 \neq 0.3$; therefore, $P(FT|TEN) \neq P(FT)$.

Based on the data, *TEN* and *FT* seem to be dependent because $P(TEN|FT) \neq P(TEN)$ and $P(FT|TEN) \neq P(FT)$.

Guided Practice: Example 2, *continued*

4. Interpret the results for $P(TEN|FT)$ and $P(FT|TEN)$.

$P(TEN|FT)$ is the probability that a student is in the tenth grade given that he prefers a field trip.

$P(FT|TEN)$ is the probability that a student prefers a field trip given that he is in the tenth grade.

The fact that *TEN* and *FT* are dependent means that being in the tenth grade affects the probability that a student prefers a field trip, and preferring a field trip affects the probability that a student is in the tenth grade.

Guided Practice: Example 2, *continued*

In this case, being in the tenth grade increases the probability that a student prefers a field trip because $P(FT|TEN) > P(FT)$. Also, preferring a field trip increases the probability that a student is in the tenth grade because $P(TEN|FT) > P(TEN)$.

$P(TEN|FT) > P(FT|TEN)$ means that it is more likely that a student is in the tenth grade given that he prefers a field trip than it is that a student prefers a field trip given that he is in the tenth grade.

Guided Practice: Example 2, *continued*

5. Compare $P(TWELVE|TS)$ and $P(TS|TWELVE)$.

$$P(TWELVE|TS) = \frac{2}{11} \approx 0.182$$

There were 11 votes for a talent show; 2 were by twelfth graders.

$$P(TS|TWELVE) = \frac{2}{12} \approx 0.167$$

There were 12 votes by twelfth graders; 2 were for a talent show.

Guided Practice: Example 2, *continued*

0.182 > 0.167; therefore, $P(TWELVE|TS) >$ $P(TS|TWELVE)$, but they are close in value. The values are approximately 18% and 17%.

Guided Practice: Example 2, *continued*

6. Determine if *TWELVE* and *TS* are independent.

Events A and B are independent events if

$P(B|A) = P(B)$ or if $P(A|B) = P(A)$.

Compare $P(TWELVE|TS)$ with $P(TWELVE)$ and $P(TS|TWELVE)$ with $P(TS)$.

Guided Practice: Example 2, *continued*

$$P(TWELVE|TS) = \frac{2}{11} \approx 0.182$$

There were 11 votes for a talent show; 2 were by twelfth graders.

$$P(TWELVE) = \frac{12}{40} = 0.3$$

There were 40 votes in all; 12 were by twelfth graders.

$0.182 \neq 0.3$; therefore, $P(TWELVE|TS) \neq P(TWELVE)$.

Guided Practice: Example 2, *continued*

$$P(TS \mid TWELVE) = \frac{2}{12} \approx 0.167$$

There were 12 votes by twelfth graders; 2 were for a talent show.

$$P(TS) = \frac{11}{40} = 0.275$$

There were 40 votes in all; 11 were for a talent show.

$0.167 \neq 0.275$; therefore, $P(TS \mid TWELVE) \neq P(TS)$.

Guided Practice: Example 2, *continued*

Based on the data, *TWELVE* and *TS* are dependent because $P(TWELVE|TS) \neq P(TWELVE)$ and $P(TS|TWELVE) \neq P(TS)$.

Guided Practice: Example 2, *continued*

7. Interpret the results for $P(TWELVE|TS)$ and $P(TS|TWELVE)$.

$P(TWELVE|TS)$ is the probability that a student is in the twelfth grade given that the student prefers a talent show.

$P(TS|TWELVE)$ is the probability that a student prefers a talent show given that the student is in the twelfth grade.

Guided Practice: Example 2, *continued*

The fact that *TWELVE* and *TS* are dependent means that being in the twelfth grade affects the probability that a student prefers a talent show, and preferring a talent show affects the probability that a student is in the twelfth grade. In this case, being in the twelfth grade decreases the probability that a student prefers a talent show because $P(TS|TWELVE)$ < $P(TS)$. And preferring a talent show decreases the probability that a student is in the twelfth grade because $P(TWELVE|TS) < P(TWELVE)$.

Guided Practice: Example 2, *continued*

$P(TWELVE|TS) > P(TS|TWELVE)$, but they are close in value, differing by only about 1%. So it is almost equally likely that a student is in the twelfth grade given that the student prefers a talent show, as it is that a student prefers a talent show given that the student is in the twelfth grade.

✔

Guided Practice: Example 2, *continued*