

## 7.6 Probability Notes

### Two-Way Table Probability Analysis (Frequency Tables)

Below is a two-way table created from the CCC Math students in Fall of 2017. The two-way table describes the relationship between whether or not a student still lives with their parents and the type of transportation they use when they go to school.

#### Contingency table results:

Rows: Do you live at home with your parents?

Columns: What type of transportation do you take to campus?

|                       | Bicycle | Carpool | Drive alone | Dropped off by someone | Public transportation | Walk | Total |
|-----------------------|---------|---------|-------------|------------------------|-----------------------|------|-------|
| Not live with parents | 0       | 7       | 58          | 1                      | 2                     | 10   | 78    |
| Lives with parents    | 1       | 23      | 205         | 17                     | 4                     | 0    | 250   |
| Total                 | 1       | 30      | 263         | 18                     | 6                     | 10   | 328   |

#### Basic Marginal Probability

Remember a percentage (proportion) is an amount out of the total. Multiply by 100% if you want to convert the proportion into a percentage. *ie: 6 boys in a class of 15.*

$$\text{percentage} = \frac{\text{amount}}{\text{total}} \cdot 100\%$$

Example 1:

Suppose we want to find the probability that a student does not live with their parents. We have a common notation when finding the probability. We would write that as

$$P(\text{does not live with parents})$$

*Probability Statement!*

This probability can be calculated by finding the amount out of the total.

$$P(\text{does not live with parents}) = \frac{\text{amount}}{\text{total}}$$

What is the amount? What is the total?

Let us start with the total. In a two way table, always use the **grand total** (sum of all the cells) unless instructed otherwise. That means you will always use the grand total unless we have a condition like restricting our answer to only people that drive alone to school. If there is no condition, always use the grand total.

What about the amount? We need the amount of people that do not live with their parents. That is actually the first row total. So our answer will be...

$$P(\text{not live with parents}) = \frac{78}{328} = 0.2378048 \approx 0.238 \rightarrow 23.8\%$$

We usually round probabilities to 3 significant figures or the thousandths place.

How can we convert this answer into a percentage?

$$\text{Multiply by } 100\%. \text{ So } 0.238 \times 100\% = 23.8\%$$

#### Contingency table results:

Rows: Do you live at home with your parents?

Columns: What type of transportation do you take to campus?

|                       | Bicycle | Carpool | Drive alone | Dropped off by someone | Public transportation | Walk | Total |
|-----------------------|---------|---------|-------------|------------------------|-----------------------|------|-------|
| Not live with parents | 0       | 7       | 58          | 1                      | 2                     | 10   | 78    |
| Lives with parents    | 1       | 23      | 205         | 17                     | 4                     | 0    | 250   |
| Total                 | 1       | 30      | 263         | 18                     | 6                     | 10   | 328   |

#### Example 2

What percentage of math students drive alone?  $\rightarrow 263$

Again, find the amount out of the total. Since there is no condition given about the student, we will use the grand total. The amount that drive alone is the total in the drive alone column.

$$P(\text{Drive Alone}) = \frac{263}{328} = 0.80182968 \approx 0.802$$

$$80.2\%$$

Joint Probabilities

Sometimes we want to find a probability when we know more than one thing about the person.

These are called Joint Probabilities. In general there are two types. "AND" = means both things have to be true about the person. "OR" means either 1 of two things are true about the person. There are no conditions (If or Given) so we will use the grand total for our total.

Two way table "AND" probabilities are a single cell out of the grand total.

**Contingency table results:**

Rows: Do you live at home with your parents?

Columns: What type of transportation do you take to campus?

|                       | Bicycle Carpool | Drive alone | Dropped off by someone | Public transportation | Walk | Total             |
|-----------------------|-----------------|-------------|------------------------|-----------------------|------|-------------------|
| Not live with parents | 0               | 58          | 1                      | 2                     | 10   | 78                |
| Lives with parents    | 1               | 205         | 17                     | 4                     | 0    | 250               |
| Total                 | 1               | 30          | 18                     | 6                     | 10   | Grand Total = 328 |

**Example 3:**

Find the probability that a student does not live with parents AND drives alone to school?

$$P(\text{not live with parents AND drives alone}) = \frac{\text{amount}}{\text{total}}$$

How many students both do not live with parents and drive alone? Both things have to be true about the person. Look for the cell where drive alone column meets the not live with parents row.

$$P(\text{not live with parents AND drives alone}) = \frac{58}{328} \approx 0.176829 \approx 0.177 \rightarrow 17.7\%$$

Two way table "OR" probabilities are more difficult because either thing could be true about the person.

**Example 4**

How would the previous problem be different if we looked for the probability of a student either not living with parents OR driving alone?

Note this will include everyone that does not live with parents (regardless of transportation status) and everyone that drives alone (regardless of parent status).

When you see "OR", you will need to add.

Circle all of the people in either the drive alone column or the not live with parents row. Add all of these to get the amount of students that do not live with parents OR drive alone.

|                       | Bicycle Carpool | Drive alone | Dropped off by someone | Public transportation | Walk | Total             |
|-----------------------|-----------------|-------------|------------------------|-----------------------|------|-------------------|
| Not live with parents | 0               | 58          | 1                      | 2                     | 10   | 78                |
| Lives with parents    | 1               | 205         | 17                     | 4                     | 0    | 250               |
| Total                 | 1               | 30          | 18                     | 6                     | 10   | Grand Total = 328 |

P (not live with parents OR drives alone) =

$$\frac{(0 + 58 + 1 + 2 + 10 + 205)}{328} = \frac{283}{328} \approx 0.8628048 \approx 0.863 \rightarrow 86.3\%$$

NOTE: Notice that we do not want to add things twice. The 58 is in both the drive alone column and the not live with parents row, but should only be added once.

NOTE: You will get the wrong answer if you add the column total (263) and the row total (78) because you will have added the 58 twice.

**Example 5:**

Find the probability that a student BOTH lives with their parents AND carpools. (Remember a single cell out of the grand total.)

$$\frac{23}{328} = 0.07012195 \approx 0.070 \rightarrow 7\%$$

**Example 6:**

What if it was an OR problem? How would it be different? Find the probability that a student EITHER lives with their parents OR carpools. (Now we have to add all students that fall in either category and do not add the same cell twice.)

$$\frac{(7 + 23 + 1 + 205 + 17 + 10)}{328} = \frac{257}{328} \approx 0.784 \rightarrow 78.4\%$$

Conditional Probabilities

Sometimes we know some prior information about a situation or person. For example, what is the probability that the L.A. Clippers basketball team will win a game against the Houston Rockets?

Suppose you found out that the Clippers 3 best players Chris Paul, Blake Griffin and Deondre Jordan all have the flu and will miss the game. Now what is the probability that the Clippers win?

This prior information, can drastically change the probability of an event. It is called a conditional probability and usually denoted with an "IF" or a "GIVEN".

|                       | Bicycle Carpool | Drive alone | Dropped off by someone | Public transportation | Walk | Total             |
|-----------------------|-----------------|-------------|------------------------|-----------------------|------|-------------------|
| Not live with parents | 0               | 7           | 58                     | 2                     | 10   | 78                |
| Lives with parents    | 1               | 23          | 205                    | 4                     | 0    | 250               |
| Total                 | 1               | 30          | 263                    | 6                     | 10   | Grand Total = 328 |

**Example 7:** Refers to a conditional probability

(if a student is dropped off by someone, what is the probability that that person lives with their parents?)

In probability theory we often write this as

$$P(\text{lives with parents} | \text{dropped off}) \text{ is the "IF" / "Given"}$$

The straight bar means "given" this is true.

The secret with conditional probabilities is to circle the row or column that has the IF or Given and then only use numbers in that row or column.

In this problem the IF was being dropped off by someone. So that restricts us to the dropped off by someone column. Circle the dropped off by someone column and use the amount and total from that column.

$$P(\text{lives w/parents} | \text{dropped off}) = \frac{17}{18} \approx 0.944 \rightarrow 94.4\%$$

|                       | Bicycle Carpool | Drive alone | Dropped off by someone | Public transportation | Walk | Total             |
|-----------------------|-----------------|-------------|------------------------|-----------------------|------|-------------------|
| Not live with parents | 0               | 7           | 58                     | 2                     | 10   | 78                |
| Lives with parents    | 1               | 23          | 205                    | 4                     | 0    | 250               |
| Total                 | 1               | 30          | 263                    | 6                     | 10   | Grand Total = 328 |

**Example 8:**

Let's try another conditional probability. Find the probability that a person walks to school if we are given that the person does not live with their parents.

$$P(\text{walks} | \text{not live with parents})$$

Remember circle the row or column with the given. Since the given is that the person does not live with parents, we should circle that row and only use the numbers in that row.

$$P(\text{walks} | \text{not live w/parents}) = \frac{10}{78} \approx 0.128 \rightarrow 12.8\%$$

Conditional Probabilities and Independence

Conditional probabilities are the key to determining if categorical variables are independent (not related) or dependent (related).

Categories are independent of each other if the condition does not matter.

Look at the following two events: L.A. Clippers basketball team winning and a new British prime minister being elected. If the events are independent then one event happening does not change the probability of the other event happening.

So if L.A. Clippers winning and a new prime minister being elected are independent,  $P(\text{L.A. Clippers win}) \approx P(\text{L.A. Clippers win} | \text{prime minister elected}) \approx P(\text{L.A. Clippers win} | \text{prime minister not elected})$

Notice the condition of prime minister did not make any difference to the probability of the L.A. Clippers winning. Since the probabilities are pretty close, this indicates that the events are independent (NOT related)

Now let's look at the example of L.A. Clippers winning and their three best players having the flu and not playing. In that case

$$P(\text{L.A. Clippers win}) \neq P(\text{L.A. Clippers win} | \text{three best players have flu})$$

≠ P(L.A. Clippers win | three best players do not have the flu)

Since the probabilities are very different, this indicates that the events are Dependent (related).

|                          | Bicycle<br>Carpool | Drive<br>alone | Dropped off by<br>someone | Public<br>transportation | Walk | Total                |
|--------------------------|--------------------|----------------|---------------------------|--------------------------|------|----------------------|
| Not live with<br>parents | 0                  | 58             | 1                         | 2                        | 10   | 78                   |
| Lives with<br>parents    | 1                  | 205            | 17                        | 4                        | 0    | 250                  |
| Total                    | 1                  | 263            | 18                        | 6                        | 10   | Grand Total =<br>328 |

**Example 9:**

Is being dropped off by someone related to (dependent) whether or not that person lives with their parents?

Look at the following probabilities. If they are close, then the events are probably independent (not related). If they are significantly different, then the events are probably dependent (related).

$$P(\text{dropped off}) = \frac{18}{328} = 0.055 \approx 5.5\%$$

$$P(\text{dropped off} | \text{lives with parents}) = \frac{17}{250} \approx 0.068 \rightarrow 6.8\%$$

$$P(\text{dropped off} | \text{NOT Live with parents}) = \frac{1}{78} \approx 0.013 \rightarrow 1.3\%$$

Since the probabilities are pretty different, this gives us some evidence that being dropped off and living with parents or not are related (dependent).