Week Four: Probability Rules

Week Four Goals

- Cross-tabulation and Two-way tables (contingency tables)
- Basic probability rules
- Addition Rules
- The Multiplication Rule
- Conditional probability
- Show that two events are statistically independent

Cross-tabulation and Two-Way Tables

If you are interested in studying two categorical variables, the data is commonly organized by CROSS-TABULATING and placed into a two-way table (in other words, a contingency table). For reference, a cross-tabulation (or crosstab) is a two- (or more) dimensional table that records the number (frequency) of respondents that have the specific characteristics described in the cells of the table. Cross-tabulation tables provide a wealth of information about the relationship between the variables.

Cross-tabulation analysis goes by several names in the research world including crosstab, contingency table, chi-square and data tabulation. Cross-tabulation is usually performed on categorical data that can be divided into mutually exclusive groups (events that can't happen at the same time).

Consider the sample data set below. It displays details about commercial transactions for four product categories (P1, P2, P3, P4). Using this raw survey data, we will be able to build a contingency table, or two-way table, summarizing the information we are interested in. Let's say we are interested in studying **payment method** (column 1) and the **product category** (column 3), we could use technology to organize the data in a two-way table. In other words, we can cross-tabulate the data.

Column 1	Column 2	Column 3	Column 4	Column 5	Column 6	Column 7
Payment Method	Coupon Applied	Product Category	Region	Price	Units	Sales
Master Card	Yes	P2	East	$19.95	2	39.90
Master Card	Yes	P3	West	$22.95	1	22.95
Master Card	No	P4	East	$19.95	1	19.95
Master Card	No	P1	North	$22.95	5	114.75
Visa	No	P1	West	$22.95	1	22.95
Visa	No	P1	East	$19.95	3	59.85
Paypal	No	P1	South	$22.95	2	45.90
Paypal	No	P1	South	$22.95	1	22.95
American Express	Yes	P2	Mid-West	$19.95	1	19.95
American Express	Yes	P2	South	$22.95	1	22.95
Visa	Yes	P2	Mid-West	$19.95	2	39.90
Paypal	Yes	P3	South	$22.95	2	45.90

Cross tabulation makes it easier to interpret data. Many studies suggest that cross tabulation is one of the most preferred methods of analyzing market research or survey data. It’s invaluable for uncovering hidden relationships in your raw data and Excel (pivot tables) and Minitab (Cross-tabulation) can do this for you. We do not need to learn HOW to cross-tabulate. We will learn to work with cross-tabulated data.

See the cross-tabulation below.

Product Category

Payment Method	P1	P2	P3	P4	Total
American Express	1	2	0	3	6
Master Card	4	6	4	5	19
Paypal	10	0	4	3	17
Visa	11	8	7	5	31
Total	26	16	15	16	73

The raw data has been counted and organized into this two-way table (contingency/pivot table).

Basic Probability Rules

Please consider visiting the Math is Fun website to learn more about probability basics.

- A probability can be written as a fraction or decimal equivalent.
- It is always a number between 0 and 1.
- If the probability of an event happening equals **ZERO**, it means there is **NO CHANCE** the event will happen.
- If the probability of an event happening equals **ONE**, it means the event is **CERTAIN to happen**.
- If the probability of an event happening equals **½**, it means there is an **even chance** of the event happening.

Probability is Just a Guide. Probability does not tell us exactly what will happen, it is just a guide.

Example 1

If you toss a coin 100 times, how many Heads will come up?
ANSWER: Probability says that heads have a ½ chance, so we can expect 50 Heads. But when we actually toss the coins, we might get 48 heads, or 55 heads ... or anything really, but in most cases, it will be a number near 50.

Complement

The complement of an event is all the **other** outcomes (**not** the ones we want). Together, the event and its complement, make all possible outcomes. The probability of an event is shown using the capital letter **P(A)**. The complement is shown by a superscript 'c' after the letter such as **$P(A^c)$** means "Probability of the complement of Event A".

Notation

P(A) means "Probability of Event A".
$P(A^c)$ means "Probability of the complement of Event A".

NOTE: The probability and its complement always add to 1.

Example 2

If the probability that it will rain tomorrow is .45. Find the probability that it will not rain.
ANSWER: $P(Rain^c) = 1 - .45 =$ **.55**

Video discussing basic probability rules from Prof. Coffey

Here is a video discussing the basic probability rules mentioned above: https://youtu.be/Qg-Zncoir3g

Here is the same video with interpreting: https://youtu.be/VHVfpRAl7f8

Examples using Basic Probability Rules

Example 3

We will use the two-way table above to answer some basic probability questions.
Round to 3 decimal places.

Product Category

Payment Method	P1	P2	P3	P4	Total
American Express	1	2	0	3	6
Master Card	4	6	4	5	19
Paypal	10	0	4	3	17
Visa	11	8	7	5	31
Total	26	16	15	16	73

This is the table total.

A. Find the probability that a product from category 2 (P2) is purchased.

Find the total number from product category 2 and divide by the table total.
Answer: P(P2) = 16/73 = **.219**

B. Find the probability that a purchase is made with Visa.

Find the total number purchased with Visa and divide by the table total.
Answer: P(Visa) = 31/73 = **.425**

C. Find the probability that a purchase is NOT made with Visa.

Find the total number purchased with Visa and divide by the table total (part B). Subtract this result from 1.
Answer: P(Visa^c) = 1 - (31/73) = 1 - .425 = **.575**

Example 4

The make-up of a British police station is shown in the two-way table below. Use this table to answer the following questions. Round to 2 decimal places.

	Male	Female	TOTAL
Constable	56	23	79
Sergeant	8	5	13
Inspector	2	4	6
Chief Inspector	1	1	2
TOTAL	67	33	100

A. What is the probability that a person working at this police station is a sergeant?

Find the total number of sergeants and divide by the table total.
ANSWER: P(sergeants) = 13/100 = **.13**

B. What is the probability that a person working at this police station is NOT a sergeant?

Subtract the answer from Part A from 1.
ANSWER: 1 - P(sergeants) = 1 - (13/100) = 1 - .13 = **.87**

Example 5

A company gathered information about whether employees were left or right-handed and in which department they worked. The results were organized in the following two-way table:

Handedness	Department 1	Department 2	Department 3
Right-handed (RH)	15	18	57
Left-handed (LH)	5	2	3

The video linked below will work through the following questions, show the notation used and how to use the shading technique with a two-way table:

1. Find the probability that an employee is left-handed.
2. Find the probability that an employee works in Department 3.
3. Find the probability that an employee does not work in Department 1.
4. Find the probability that an employee works in Departments 2 and 3.

Video showing Example 5 from Prof. Coffey

Here is a video discussing Example 5: https://youtu.be/Qg-Zncoir3g

Here is the same video with interpreting: https://youtu.be/0VGRUEcfiOI

Addition Rules

Addition Rule 1

When two events A and B are mutually exclusive, the probability that A or B will occur is the sum of the probability of each event: P(A or B) = P(A) + P(B)

Instead of writing OR, we often us the 'in union with' symbol: U
The formula is then rewritten as:

$$P(A \cup B) = P(A) + P(B)$$

This is read: The probability of A or B equals the probability of A added to the probability of B.
HINT: When you read 'OR' think 'ADD'. We study the combined outcomes of both events.

Example 6

A single 6-sided die is rolled. What is the probability of rolling a 2 or a 5?
There is one side representing a '2'. There are 6 sides. There is one side representing a '5'.
Answer: P(2) = 1/6 and the P(5) = 1/6. Rolling a 2 or a 5 are mutually exclusive; they both CANNOT happen.
Therefore, P(2 or 5) = P(2 U 5) = 1/6 + 1/6 = 2/6 which reduces to 1/3.

Addition Rule 2

When two events A and B are non-mutually exclusive, the probability that A or B will occur is P(A or B) = P(A) + P(B) – P(A and B).

In the rule above, P(A and B) refers to the overlap of the two events. Since we will be using two-tables to answer our probability questions, we will not need to use the addition rule 2 explicitly. Instead, we will use the idea of overlap and shade the cells that represent our event. See examples using the shading technique.

Example 7

The following two-way table represents data collected from a British police station. We have the rank and gender for the members of this police station.

	Male	Female	Total
Constable	56	23	79
Sergeant	8	5	13
Inspector	2	4	6
Chief inspector	1	1	2
Total	67	33	100

A. Find the probability that an employee at the police station is an inspector **or** a female.

Answer: Since **there is an overlap** of females and inspectors, we will have to be careful not to count either group twice. This can be done visually by shading those cells that represent inspectors or females. Add up the shaded cells and you get 35. Divide by the table total. P(inspector or female) = 35/100 = **.35.**

	Male	Female	Total
Constable	56	23	79
Sergeant	8	5	13
Inspector	2	4	6
Chief inspector	1	1	2
Total	67	33	100

I have shaded those cells that are inspectors and those that are female.

NOTE: The table shading method takes much less time than using Addition Rule 2:
P(Inspector ∪ Female) = P(Inspector) + P(Female) – P(Inspector and Female)
P(Inspector ∪ Female) = (6/100) + (33/100) - (4/100) = 35/100
Either way, however, we get to the same answer.

B. What is the probability that an employee at this police station is a constable OR an inspector?

	Male	Female	Total
Constable	56	23	79
Sergeant	8	5	13
Inspector	2	4	6
Chief inspector	1	1	2
Total	67	33	100

Answer: Since **there is no overlap** of constable and inspectors, we have any easier problem. Shade those cells that represent constables or inspectors. Add up the shaded cells and you get 79 + 6 = 85. Divide by the table total. P(constable ∪ inspector) = 85/100 = **.85.**

	Male	Female	Total
Constable	56	23	79
Sergeant	8	5	13
Inspector	2	4	6
Chief inspector	1	1	2
Total	67	33	100

C. What is the probability that an employee at this police station is a male or constable?

	Male	Female	Total
Constable	56	23	79
Sergeant	8	5	13
Inspector	2	4	6
Chief inspector	1	1	2
Total	67	33	100

Answer: Since **there is an overlap** of constable and males, we have to be sure not to count the same responses twice. Shade those cells that represent all males. Also shade all the constables. Add up the shaded cells and you get 67 + 23 = 90. Divide by the table total. P(male $\cup$ constable) = 90/100 = **.90.**

	Male	Female	Total
Constable	56	23	79
Sergeant	8	5	13
Inspector	2	4	6
Chief inspector	1	1	2
Total	67	33	100

Adapted from this website.

Example 8

A company gathered information about whether employees were left or right-handed and in which department they worked. The results were organized in the following two-way table:

Handedness	Department 1	Department 2	Department 3
Right-handed (RH)	15	18	57
Left-handed (LH)	5	2	3

The video linked below will work through the following questions, show the notation used and how to use the shading technique with a two-way table:

1. Find the probability that an employee works in Department 1 or is left-handed.
ANSWER: .25

2. Find the probability that an employee is not left-handed or works in Department 2.
ANSWER: .92

Video showing Example 8 from Prof. Coffey

Here is a video discussing Example 8: https://youtu.be/n2Ii69PYlaA

Here is the same video with interpreting: https://youtu.be/BDHDCbmzNcE

Multiplication Rule

When two events A and B are independent, the probability of both occurring is P(A and B) = P(A) · P(B).

Instead of writing 'AND', probability notation uses the 'intersection' symbol: $\cap$

Independent Events

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

This is read: The probability of A and B equals the probability of A multiplied by the probability of B.

HINT: When you read 'AND think 'MULTIPLY'. We study only the outcomes that are common to both events.

Independent vs. Dependent Events

- Two events are **dependent** if the outcome or occurrence of the first affects the outcome or occurrence of the second so that the probability is changed.
- Two events are **independent** if the outcome or occurrence of the first DOES NOT affect the outcome or occurrence of the second.

NOTE: there is another multiplication rule for dependent events. This rule uses conditional probability and that will be covered in the next section.

Example 9

Using the police station two-way table again…

	Male	Female	Total
Constable	56	23	79
Sergeant	8	5	13
Inspector	2	4	6
Chief inspector	1	1	2
Total	67	33	100

A. What is the probability that an employee is a sergeant and female?

Answer: Find the cell that represents the intersection of females and sergeants. There are 5 observations. P(female $\cap$ sergeant) = 5/100.

	Male	Female	Total
Constable	56	23	79
Sergeant	8	5	13
Inspector	2	4	6
Chief inspector	1	1	2
Total	67	33	100

I only shaded the intersection of females and sergeants.

B. What is the probability that an employee is NOT a constable and is female?

Answer: Find the cell that represents the intersection of females and not constables (i.e. sergeants, inspectors and chief inspectors). Shade all of the female sergeants, female inspectors and female chief inspectors.
P(not constable ∩ female) = 10/100. I could have also written this: P(constable^c ∩ female).

	Male	Female	Total
Constable	56	23	79
Sergeant	8	5	13
Inspector	2	4	6
Chief inspector	1	1	2
Total	67	33	100

Example 10

A company gathered information about whether employees were left or right-handed and in which department they worked. The results were organized in the following two-way table:

Handedness	Department 1	Department 2	Department 3
Right-handed (RH)	15	18	57
Left-handed (LH)	5	2	3

The video linked below will work through the following question: Find the probability that an employee works in Department 2 **and** is left-handed.
ANSWER: .02

Example 11

The video linked below will also work through the following question: Assuming the probability that a baby is born a 'boy' on any given day is .48, what is the probability that a baby boy is born on 3 consecutive days?
ANSWER: .11

Video showing Examples 10 and 11 from Prof. Coffey

Here is a video discussing Examples 10 and 11: https://youtu.be/LgfeiE5uOa8

Here is the same video with interpreting: https://youtu.be/qbwGZnBzFL0

Practice with Addition and Multiplication Rules

Example 12

On the morning of April 10, 1912, the Titanic set sail from Southampton, England on its maiden voyage to New York City. On April 14, just four days into the trip, the Titanic struck an iceberg and, in just over 2 hours, the "unsinkable" Titanic was gone with a loss of 1,490 lives. The Titanic had been outfitted with 20 lifeboats, enough to hold about 54% of the passengers.

The table below gives a count of the survivors and victims of the Titanic sinking. The data are classified by type of passenger (first class, second class, third class, crew member) and survival status (survived or died). The data for this analysis were compiled from the initial casualty estimates, as reported in the Official Board of Trade Inquiry Report, written originally in 1912.

	Survived	Died
First Class	203	122
Second Class	118	167
Third Class	178	528
Crew Class	212	673

First, knowing the totals is helpful, especially the table total. The column and row totals are calculated.

	Survived	Died	Total
First Class	203	122	325
Second Class	118	167	285
Third Class	178	528	706
Crew Class	212	673	885
Total	711	1490	2201

Using the two-way table with totals and shading technique, answer the following questions. The answers are found below.

A. What is the probability that a randomly selected passenger survived?

B. What is the probability that a passenger survived and was in second class?

C. What is the probability that a passenger survived and was in third class or crew class?

D. What is the probability that a passenger died or was in first class or second class?

SOLUTIONS to Example 12

A. What is the probability that a randomly selected passenger survived?

	Survived	Died	Total
First Class	203	122	325
Second Class	118	167	285
Third Class	178	528	706
Crew Class	212	673	885
Total	711	1490	2201

$P(\text{surv}) = 711/2201$

$= .323$

B. What is the probability that a passenger survived and was in second class?

	Survived	Died	Total
First Class	203	122	325
Second Class	118	167	285
Third Class	178	528	706
Crew Class	212	673	885
Total	711	1490	2201

$P(\text{surv} \cap 2^{nd}) = \frac{118}{2201}$

$= .054$

C. What is the probability that a passenger survived and was in third class or crew class?

	Survived	Died	Total
First Class	203	122	325
Second Class	118	167	285
Third Class	178	528	706
Crew Class	212	673	885
Total	711	1490	2201

$P(\text{Survived} \cap 3^{rd} \text{ or Crew}) =$

$= \frac{390}{2201}$

$= .177$

D. What is the probability that a passenger died or was in first class or second class?

Looking at the shading, you can see that this is the complement of Part C. You can use complements to answer it or use my method in the image below.

Using complements: 1 - .177 = **.823**

	Survived	Died	Total
First Class	203	122	325
Second Class	118	167	285
Third Class	178	528	706
Crew Class	212	673	885
Total	711	1490	2201

$P(\text{died OR } 1^{st} \text{ OR } 2^{nd}) =$

$= \frac{1811}{2201}$

$= .823$

Conditional Probability

The **conditional probability** of an event A is the probability that the event will occur given the knowledge that an event B has already occurred. This probability is written P(A | B), notation for the probability of A given B. In the case where events A and B are independent (where event B has no effect on the probability of event A), the conditional probability of event A given event B is simply the probability of event A, that is P(A).
In other words, if events A and B are independent, P(A | B) = P(A)...because B had no effect on A (they are independent!)

If events A and B are **NOT independent**, the conditional probability P(A | B) is:

Probability of event A occured and event B occured

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Probability of event A given B has occured

Probability of event B

NOTE: If we were to rearrange this, it would be the multiplication rule for dependent events:

Dependent Events

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

The order of A and B only matters in that the Event that is the restriction must match the Event that we find the probability.

We will answer conditional probability questions using the two-way table and will not need to use the formula. Instead, we can use a shading technique.
You will know to use the CONDITIONAL PROBABILITY when the problem says something similar to:

- Find the probability of M given that S has happened: P(M | S)
- Find the probability of M restricted to S: P(M | S)
- Find the probability of M among/within S: P(M | S)

In each equivalent statement above, the 'M' event is the event we are interested in and the 'S' event is the condition that must be met.

Example 13

A. Find the probability that a person survived ***when restricted to*** those in first class?

ANSWER: Find the number the total number that were in first class---this is what we are restricted to and this is what I have put a red box around. Now, find the number that survived when restricted only to those that were in first class. This is what I have shaded in yellow.
P(Survived / First) = 203/325 = .625

	Survived	Died	Total
First Class	203	122	325
Second Class	118	167	285
Third Class	178	528	706
Crew Class	212	673	885
Total	711	1490	2201

B. Find the probability that a person was in first class when restricted to those that survived.

ANSWER: Find the number the total number that survived---this is what we are restricted to and this is what I have put a red box around. Now, find the number that were in first class when restricted only to those that survived. This is what I have shaded in yellow.
P(First / Survived) = 203/711 = .289

	Survived	Died	Total
First Class	203	122	325
Second Class	118	167	285
Third Class	178	528	706
Crew Class	212	673	885
Total	711	1490	2201

C. Find the probability that a person died among those that were in the Crew Class.
ANSWER: P(Died / Crew) = 673/885 = .76

	Survived	Died	Total
First Class	203	122	325
Second Class	118	167	285
Third Class	178	528	706
Crew Class	212	673	885
Total	711	1490	2201

Video showing Example 13 parts A and C from Prof. Coffey

Here is a video discussing Example 13: https://youtu.be/K23Wfb3vQZI

Here is the same video with interpreting: https://youtu.be/ahklkny-UmI

Show that Two Events are Independent

Now that we have learned how to answer conditional probability questions, we can use the concept to show whether or not two events are statistically independent.

We can show that two events are statistically independent if the "condition" of one event does not change the probability of the other event happening. In other words, If P(A / B) = P(A) OR P(B / A) = P(B), then A and B are independent.

Example 14

Using the Titanic data once more, we can determine if a passenger surviving has nothing to do with whether they were in First Class. In other words, we can determine if surviving and first class are statistically independent. Here is a video explanation:

Video showing Example 14 from Prof. Coffey

Here is a video discussing Example 14: https://youtu.be/l79dafZ4Exo

Here is the same video with interpreting: https://youtu.be/aJEOGHxoHuA

Example 15

Are the events being a "constable' and being 'male' independent?

	Male	Female	Total
Constable	56	23	79
Sergeant	8	5	13
Inspector	2	4	6
Chief inspector	1	1	2
Total	67	33	100

Answer: Find the P(constable) and see if it is equal to P(constable / male).
P(constable) = 79/100 = .79 P(constable / male) = 56/67 = .836
No, according to the definition of independence, being a constable and being a male are not independent of one another.

NOTE: It is equally correct to find the P(male) and see if it is equal to P(male / constable).
P(male) = 67/100 = .67 P(male / constable) = 56/79 = .709
Same result...male and constable are NOT independent.