Frequency Distribution

What Is Frequency Distribution?

Frequency Distribution

A **frequency** of a value is the number of times it occurs in a dataset. A **frequency distribution** is the number of times each variable occurs in a dataset.

Types of Frequency Distributions

- **Ungrouped frequency distribution**: the number of observations of each output. It's usable for *categorical data*.

Types of Frequency Distributions

- **Ungrouped frequency distribution**: the number of observations of each output. It's usable for *categorical data*.
- **Grouped frequency distribution**: the number of observations of each **class interval** of a variable. Useful for *quantitative data*.

Types of Frequency Distributions

- **Ungrouped frequency distribution**: the number of observations of each output. It's usable for *categorical data*.
- **Grouped frequency distribution**: the number of observations of each **class interval** of a variable. Useful for *quantitative data*.
- **Relative frequency distribution**: the proportion of each value or class interval of a variable. Useful for any type of data **if we care about comparing frequencies** rather than amounts.

Types of Frequency Distributions

- **Ungrouped frequency distribution**: the number of observations of each output. It's usable for *categorical data*.
- **Grouped frequency distribution**: the number of observations of each **class interval** of a variable. Useful for *quantitative data*.
- **Relative frequency distribution**: the proportion of each value or class interval of a variable. Useful for any type of data **if we care about comparing frequencies** rather than amounts.
- **Cumulative frequency distribution**: the sum of frequencies less than or equal to each value or class interval of a variable. Useful when we want to understand how often observations fall below certain values.

Categorical vs Quantitative Data

Quantitative Data represent real amounts that can be added, subtracted etc.

Categorical vs Quantitative Data

Quantitative Data represent real amounts that can be added, subtracted etc.

- Can be discrete or continuous:

Categorical vs Quantitative Data

Quantitative Data represent real amounts that can be added, subtracted etc.

- Can be discrete or continuous:
  - **Discrete data** represents counts of individual items like number of students in a class.

Categorical vs Quantitative Data

Quantitative Data represent real amounts that can be added, subtracted etc.

- Can be discrete or continuous:
  - **Discrete data** represents counts of individual items like number of students in a class.
  - **Continuous data** represents measurements of uncountable values like density, volume or time.

Categorical vs Quantitative Data

Categorical Data represents groupings. They can be recorded as numbers but the numbers represent categories and not actual amounts.

Categorical vs Quantitative Data

Categorical Data represents groupings. They can be recorded as numbers but the numbers represent categories and not actual amounts.

- Can be binary, nominal or ordinal:

Categorical vs Quantitative Data

Categorical Data represents groupings. They can be recorded as numbers but the numbers represent categories and not actual amounts.

- Can be binary, nominal or ordinal:
  - **Binary data** represents yes or no outcomes like coin flips or win/loss situations.

Categorical vs Quantitative Data

Categorical Data represents groupings. They can be recorded as numbers but the numbers represent categories and not actual amounts.

- Can be binary, nominal or ordinal:
  - **Binary data** represents yes or no outcomes like coin flips or win/loss situations.
  - **Nominal data** represents groups without rank or order between them – like the names of species or colours.

Categorical vs Quantitative Data

Categorical Data represents groupings. They can be recorded as numbers but the numbers represent categories and not actual amounts.

- Can be binary, nominal or ordinal:
  - **Binary data** represents yes or no outcomes like coin flips or win/loss situations.
  - **Nominal data** represents groups without rank or order between them – like the names of species or colours.
  - **Ordinal data** represents groups that are ranked – like finishing place in a race.

Ungrouped Frequency Data – Table

1. Create a table with one column for inputs and as many columns as there are output with a row for each input.

Ungrouped Frequency Data – Table

1. Create a table with one column for inputs and as many columns as there are output with a row for each input.
    - For **ordinal** variables, the values should be ordered from smallest to largest.

Ungrouped Frequency Data – Table

1. Create a table with one column for inputs and as many columns as there are output with a row for each input.
    - For **ordinal** variables, the values should be ordered from smallest to largest.
    - For **nominal** variables, the rows can be ordered arbitrarily.
2. Count the **frequencies**.

Ungrouped Frequency Data – Example 1

A gardener sets up a bird feeder in his backyard. He wishes to know which type of bird species visit the feeder the most.

Ungrouped Frequency Data – Example 1

A gardener sets up a bird feeder in his backyard. He wishes to know which type of bird species visit the feeder the most. His observations are in the following table:

Species	Frequency
Chickadee	3
Dove	1
Finch	4
Grackle	2
Sparrow	4
Starling	2

Ungrouped Frequency Data – Example 2

We observe how many times a specific type of tram (based on age) stops at a chosen station each day.

Ungrouped Frequency Data – Example 2

We observe how many times a specific type of tram (based on age) stops at a chosen station each day.
This experiment may yield a table like this:

Type	Frequency
1990	6
1996	11
2005	3
2017	5

Grouped Frequency Data – Table

1. Divide the variables into **class intervals**. There's no 'best choice' for the width of a class interval. Different choices convey different meanings.

Grouped Frequency Data – Table

1. Divide the variables into **class intervals**. There's no 'best choice' for the width of a class interval. Different choices convey different meanings.
    - Calculate the **range** = highest value − lowest value.

Grouped Frequency Data – Table

1. Divide the variables into **class intervals**. There's no 'best choice' for the width of a class interval. Different choices convey different meanings.
   - Calculate the **range** = highest value $-$ lowest value.
   - Decide on the **class interval width**. **ALWAYS THINK** about that the best width should be! But, if you can't decide, a rule of thumb is the width

$$\text{width} = \frac{\text{range}}{\sqrt{\text{number of inputs}}}.$$

   It is typically beneficial to round this value to an integer.

Grouped Frequency Data – Table

1. Divide the variables into **class intervals**. There's no 'best choice' for the width of a class interval. Different choices convey different meanings.
   - Calculate the **range** = highest value $-$ lowest value.
   - Decide on the **class interval width**. **ALWAYS THINK** about that the best width should be! But, if you can't decide, a rule of thumb is the width

$$\text{width} = \frac{\text{range}}{\sqrt{\text{number of inputs}}}.$$

     It is typically beneficial to round this value to an integer.
   - Calculate the **class intervals**. Each interval is of the form [lower limit, lower limit $+$ width). Simply divide the outputs into these intervals.

Grouped Frequency Data – Table

2. Create a **table** with columns for inputs and each output and as many rows as there are class intervals.

Grouped Frequency Data – Table

2. Create a **table** with columns for inputs and each output and as many rows as there are class intervals.
3. Count the **frequencies**.

Grouped Frequency Data – Example

In sociological surveys, you typically want to find the distribution of respondents by age.

Grouped Frequency Data – Example

In sociological surveys, you typically want to find the distribution of respondents by age. Let's say a survey had 20 respondents. Their ages go like this:

52, 34, 32, 29, 63, 40, 46, 54, 36, 36, 24, 19, 45, 20, 28, 29, 38, 33, 49, 37.

Grouped Frequency Data – Example

In sociological surveys, you typically want to find the distribution of respondents by age. Let's say a survey had 20 respondents. Their ages go like this:

$$52, 34, 32, 29, 63, 40, 46, 54, 36, 36, 24, 19, 45, 20, 28, 29, 38, 33, 49, 37.$$

We calculate the range as highest $-$ lowest $= 63 - 19 = 44$.
We calculate the interval width as

$$\text{width} = \frac{\text{range}}{\sqrt{\text{sample size}}} = \frac{44}{\sqrt{20}} = 9.84,$$

and round it up to 10.

Grouped Frequency Data – Example

In sociological surveys, you typically want to find the distribution of respondents by age. Let's say a survey had 20 respondents. Their ages go like this:

$$52, 34, 32, 29, 63, 40, 46, 54, 36, 36, 24, 19, 45, 20, 28, 29, 38, 33, 49, 37.$$

We calculate the range as highest $-$ lowest $= 63 - 19 = 44$.
We calculate the interval width as

$$\text{width} = \frac{\text{range}}{\sqrt{\text{sample size}}} = \frac{44}{\sqrt{20}} = 9.84,$$

and round it up to 10.
Therefore, we have the following intervals

$$[19, 29), [29, 39), [39, 49), [49, 59), [59, 69).$$

Grouped Frequency Data – Example

Counting the numbers of outputs falling into each of those intervals gives the table:

Age	Frequency
19 – 28	4
29 – 38	9
39 – 48	3
49 – 58	3
59 – 68	1

Relative Frequency Data – Table

1. Simply create a grouped or ungrouped frequency table.

Relative Frequency Data – Table

1. Simply create a grouped or ungrouped frequency table.
2. To each output add another column to represent **relative frequencies**.

Relative Frequency Data – Example

In our gardener example, the relative frequency table would look like this:

Species	Frequency	Relative Frequency
Chickadee	3	$\frac{3}{3+1+4+2+4+2} = 0.19$
Dove	1	0.06
Finch	4	0.25
Grackle	2	0.13
Sparrow	4	0.25
Starling	2	0.13

Cumulative Frequency Data – Table

1. Create an ungrouped or grouped frequency table for **an ordinal or quantitative variable**. Cumulative frequencies make no sense for nominal variables because they're not ordered.

Cumulative Frequency Data – Table

1. Create an ungrouped or grouped frequency table for **an ordinal or quantitative variable**. Cumulative frequencies make no sense for nominal variables because they're not ordered.
2. Add another column for each output with **cumulative frequency**. The cumulative frequency is the number of observations less than or equal to a certain value or class interval.

Cumulative Frequency Data – Example

Going back to our example of a sociological survey. The cumulative frequency table of the age of survey participants would look like this:

Age	Frequency	Cumulative Frequency
19 – 28	4	4
29 – 38	9	9 + 4 = 13
39 – 48	3	9 + 4 + 3 = 16
49 – 58	3	19
59 – 68	1	20

1

Graphing Frequency Distributions

Pie Charts

Pie charts can be used to graph **relative** frequency distributions of **nominal variables**.

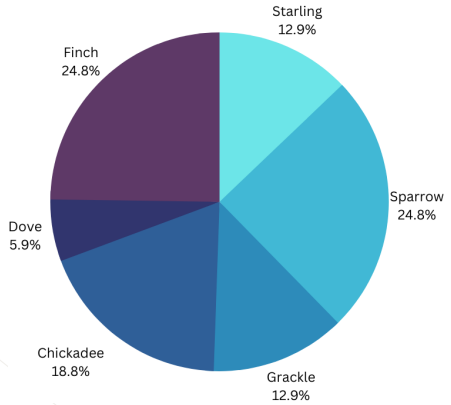
Pie Charts

Pie charts can be used to graph **relative** frequency distributions of **nominal variables**.
In our gardener example, we had the following relative frequency table of bird species:

Species	Frequency	Relative Frequency
Chickadee	3	$\frac{3}{3+1+4+2+4+2} = 0.19$
Dove	1	0.06
Finch	4	0.25
Grackle	2	0.13
Sparrow	4	0.25
Starling	2	0.13

Pie Charts

This table can be organized into a pie chart like this:



Bar Chart

Bar charts can be used to visualize the **frequency** or **relative frequency** of **categorical data** (both **nominal** and **ordinal**).

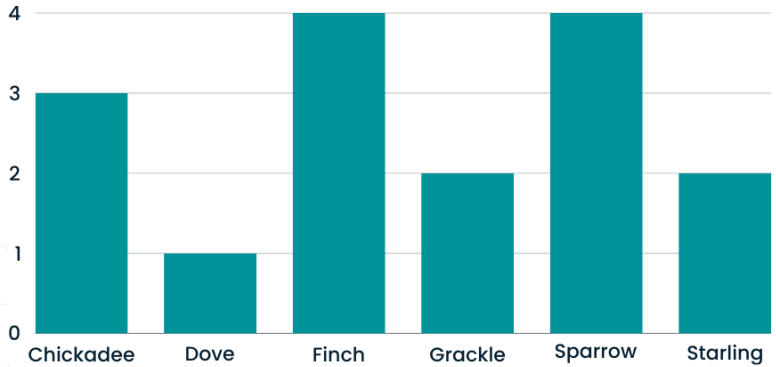
Bar Chart

Bar charts can be used to visualize the **frequency** or **relative frequency** of **categorical data** (both **nominal** and **ordinal**). The table of frequency distribution of bird species in our gardener example looked like this:

Species	Frequency
Chickadee	3
Dove	1
Finch	4
Grackle	2
Sparrow	4
Starling	2

Bar Chart

We can organize this data into a bar chart:



Histogram

A histogram looks similar to a bar chart and can be used to visualize the **frequency** or **relative frequency** of a **quantitative variable**.

Histogram

A histogram looks similar to a bar chart and can be used to visualize the **frequency** or **relative frequency** of a **quantitative variable**.
Recall the table from our sociological survey example:

Age	Frequency
19 – 28	4
29 – 38	9
39 – 48	3
49 – 58	3
59 – 68	1

Histogram

The values can be organized into the following histogram: