Representing categorical variable with tables

A survey records students' favourite type of music and presents the results in a table showing counts for Pop, Rock, Hip-Hop, and Classical. What type of table is being used?

A. Scatter plot table
B. Frequency table
C. Line graph table
D. Regression table

In a two-way table showing gender (male/female) and preferred sport (football/basketball/tennis), what is a "marginal frequency"?

A. The count in each interior cell of the table
B. The total for each row or column
C. The difference between categories
D. The percentage of the largest category

Representing categorical variable with tables

- A survey records students' favourite type of music and presents the results in a table showing counts for Pop, Rock, Hip-Hop, and Classical. What type of table is being used?
- A. Scatter plot table
  B. Frequency table
  C. Line graph table
  D. Regression table

B

-

- In a two-way table showing gender (male/female) and preferred sport (football/basketball/tennis), what is a "marginal frequency"?
- A. The count in each interior cell of the table
  B. The total for each row or column
  C. The difference between categories
  D. The percentage of the largest category

B

- Categorical variable — 分类变量
- Frequency table — 频数表
- Relative frequency — 相对频率
- Two-way table — 双向表
- Row total — 行总计
- Column total — 列总计
- Joint frequency — 联合频数
- Marginal frequency — 边际频数
- Conditional frequency — 条件频数
- Percentage — 百分比

- Categorical variables classify data into groups or categories.
    - **Frequency tables** show the count in each category.
    - **Relative frequency tables** show proportions or percentages.
- Tables make it easy to compare categories clearly and efficiently.
    - Categories should be **mutually exclusive and exhaustive**.

Political Affiliation	Frequency
Democratic	12
Republican	10
Independent	3

Frequency Table for the Political Affiliations.

Frequency Table:

Colour	Tally marks	Frequency
Black	|	1
Blue	𝍸	5
Pink	||	2
White	||||	4
		Total = 12

T/F. A frequency table always shows the exact values of all data points in a dataset.

Relative Frequency Table:

Number of tosses	Number of heads	Relative frequency
5	4	$\frac{4}{5} = 0.80$
10	6	$\frac{6}{10} = 0.60$
50	23	$\frac{23}{50} = 0.46$
100	49	$\frac{49}{100} = 0.49$

T/F A relative frequency table shows the proportion or percentage of the total data that falls into each category.

Cumulative Frequency Table:

Grade	Frequency	Cumulative Frequency
1	2	2
2	2	2 + 2 = 4
3	4	4 +4 = 8
4	2	2 + 8 = 10
5	4	4 + 10 = 14
6	2	2 + 14 = 16

T/F. A cumulative frequency table shows the running total of frequencies up to each category or class in the dataset.

Two-Way (Contingency) Table:

Contingency Table			
	Boy	Girl	Sum
like Snickers	43	30	73
doesn't like Snickers	8	19	27
Sum	51	49	100

T/F A contingency table displays the relationship between two categorical variables by showing the frequency of all possible combinations of the variables.

A school collects data from 200 students about their **mode of transport to school** and **year group**. The results are shown in the two-way table below:

	Bus	**Car**	**Walk**	**Total**
Year 9	30	20	10	60
Year 10	25	35	20	80
Year 11	15	25	20	60
Total	70	80	50	200

(a) Identify the categorical variables in this study.

(b) What is the marginal frequency of students who travel by car?

(c) Calculate the conditional frequency (as a proportion) of Year 10 students who walk to school.

(d) Which year group is most likely to walk to school? Justify your answer using appropriate calculations.

(e) Describe one pattern or relationship between year group and mode of transport.

(a) Year group, mode of transport

(b) 80

(c) $20/80 = 0.25$

(d) Year 11: $20/60 \approx 0.33$ (highest proportion walking)

(e) Older students (Year 11) are more likely to walk compared to younger students.

Stretch and Challenge

Stretch & Challenge Questions – Representing Categorical Variables with Tables

1. A survey records students' preferred learning style (visual, auditory, kinesthetic) and year group. The data is shown in a two-way table. Explain why comparing **conditional distributions (row percentages)** is more useful than comparing raw frequencies when determining association.

---

2. A researcher claims that if two categorical variables have similar conditional distributions across all groups, then they must be independent. Do you agree? Justify your answer.

---

3. In a two-way table, one category has a very large marginal frequency compared to others. Explain how this could affect the interpretation of association between the two variables.

---

4. A dataset shows a strong association between "ice cream sales" and "number of drowning incidents." Explain why a two-way table alone cannot be used to conclude a causal relationship, and suggest a possible lurking variable.

- Conditional distributions allow fair comparison by accounting for different group sizes, making it easier to detect patterns/association.
- Disagree. Similar distributions suggest independence but do not prove it exactly due to possible sampling variation.
- It can bias interpretation because large groups may dominate totals, hiding patterns in smaller groups.
- Correlation does not imply causation. A lurking variable (e.g., temperature/season) affects both ice cream sales and swimming activity.

Exploring Two-Variable Categorical Data (Associations in Two-Way Tables)

- Two-way table — 双向表
- Categorical variable — 分类变量
- Association — 关联关系
- Independence — 独立性
- Conditional distribution — 条件分布
- Marginal distribution — 边际分布
- Joint frequency — 联合频数
- Relative frequency — 相对频率
- Row percentage — 行百分比
- Column percentage — 列百分比

- A **two-way table** is used to display the relationship between **two categorical variables**.
- Each cell shows a **joint frequency** (count for a combination of categories).
- **Marginal frequencies** are the totals for each row or column.
- **Conditional distributions** show the breakdown of one variable **given a specific category of the other variable**.
- Conditional distributions are often expressed as **relative frequencies or percentages**.
- Comparing **conditional distributions** helps determine whether variables are **associated**.
- If conditional distributions are similar across groups, the variables may be **independent**.
- If distributions differ noticeably, there is likely an **association** between variables.
- **Row or column percentages** are used to compare groups fairly.
- The goal is often to determine whether there is **evidence of a relationship (not causation)** between variables.

A school surveys 150 students to study the relationship between **year group** and **preferred study location.** The results are shown below:

	Library	**Home**	**Café**	**Total**
Year 9	20	25	5	50
Year 10	15	20	15	50
Year 11	25	15	10	50
Total	60	60	30	150

(a) Identify the two categorical variables in this study.

(b) Find the marginal frequency of students who prefer studying at home.

(c) Calculate the conditional distribution (as proportions) of Year 10 students across study locations.

(d) Which year group is most likely to study in a café? Show calculations to support your answer.

(e) Describe whether there appears to be an association between year group and study location. Justify your answer using the table.

(a) Year group, preferred study location

(b) 60

(c) Year 10:
Library = 15/50 = 0.30
Home = 20/50 = 0.40
Café = 15/50 = 0.30

(d) Year 10: 15/50 = 0.30 (highest proportion for café)

(e) There is some association because Year 10 has a higher proportion choosing cafés compared to other years.

True or False

- A two-way table is used to display the relationship between two quantitative variables.

A two-way table is used for **two categorical variables**, not quantitative variables. It shows counts (or proportions) for combinations of categories, not numerical relationships like mean vs height.

True or False

- Relative frequencies are often expressed as percentages.

TRUE

True or False

- Association means one variable causes the other.

If two variables are associated, it only means they are related or show a pattern together. It does **not prove that one causes the other**—other factors may be involved or it may be coincidental.