Section 3.1.11 – Two Way Frequency Tables & Segmented Bar Charts

VCAA "Dot Points"
Investigating data distributions, including:

- *review of representation, display and description of the distributions of categorical variables: data tables, two-way frequency tables and their associated segmented bar charts*

Two Way Frequency Table

Two way frequency tables are used to provide comparisons between **two categorical variables**. Take for example, the favourite leisure activities for 50 adults. 20 men and 30 women were surveyed and their responses are displayed below in a **two way frequency table**:

Leisure Activity	Men	Women	Total
Dance	2	15	17
Sport	12	12	24
TV	6	3	9
Total	20	30	50

Marginal distributions are the sums of the **rows** or the **columns** and are found in the **margins** of the table.

Leisure Activity	Men	Women	Total
Dance	2	15	17
Sport	12	12	24
TV	6	3	9
Total	20	30	50

Conditional distribution are the numbers that make up the **sub population**. In this case the values under the headings of men and women are the conditional distributions. For example the conditional distributions for Men in the above example are shown in **red**.

Leisure Activity	Men	Women	Total
Dance	2	15	17
Sport	12	12	24
TV	6	3	9
Total	20	30	50

The two categorical variables considered in this scenario are: **leisure activity** and **gender**. In this example we are trying to identify if the leisure activities are dependent upon the participants gender.

NB: Both the horizontal and vertical totals add to 50.

In most examples

- the **explanatory variable** (independent variable) will be located in the **columns** of the table (in this case the respondent's gender is referred to as the explanatory variable)

- the **response variable** (dependent variable) will be located in the **rows** of the table (in this case leisure activity is referred to as the response variable)

In this example it would appear that both Sport and TV experience similar participation by both genders. Dance appears to be the only leisure activity where there is a significant gender discrepancy, whereby there are far more women participants than men.

For a more conclusive comparison we need to consider the make-up of the cohort being surveyed. In which case we are better to construct a **two way percentage frequency table**.

Two Way Percentage Frequency Table

In the previous example we were comparing the leisure activity of both genders. In order to compare responses fairly we must first consider the fact that more women were surveyed than men. Accordingly, a better measure of comparison is made using a **two way percentage frequency table**.

A two way percentage frequency table uses **percentage** as a measure of comparison.

Two way frequency table

Leisure Activity	Men	Women
Dance	2/20	15/30
Sport	12/20	12/30
TV	6/20	3/30
Total	20	30

Two way percentage frequency table

Leisure Activity	Men	Women
Dance	10%	50%
Sport	60%	40%
TV	30%	10%
Total	100%	100%

A two way percentage frequency table provides information, from which we can make **accurate comparisons**. In this example we can state that:

- 60% of men prefer Sport as a leisure activity, whilst only 40% of women do
- 50% of women prefer Dance as a leisure activity, whilst only 10% of men do

Alternatively, you could also be asked to calculate percentages from the table rows, rather than columns.

For example: Of those surveyed who enjoy Dance as their preferred leisure activity, what percentage were women?

Leisure Activity	Men	Women	Total
Dance	2	15	17
Sport	12	12	24
TV	6	3	9

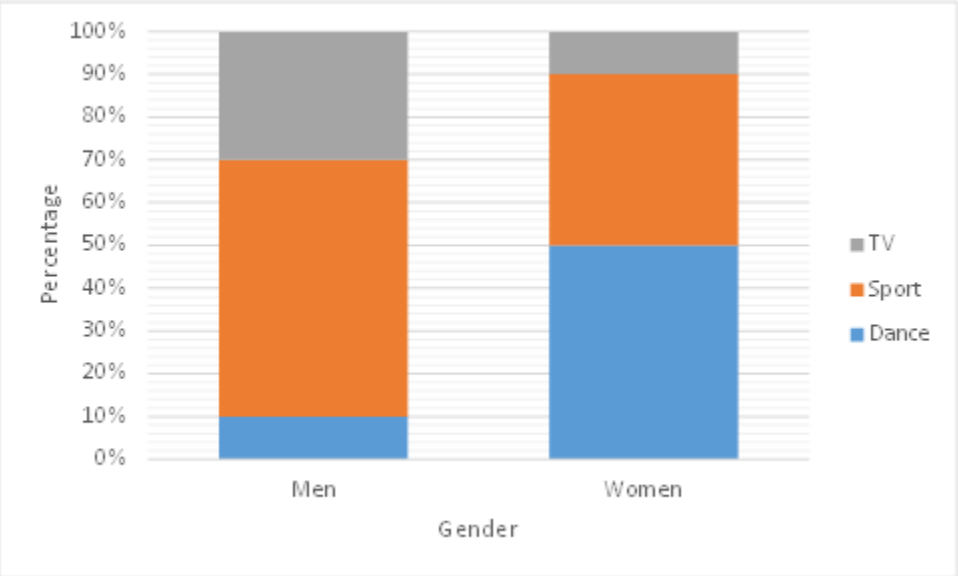
Leisure Activity	Men	Women	Total
Dance	2/17	15/17	17
Sport	12/24	12/24	24
TV	6/9	3/9	9

Leisure Activity	Men	Women	Total
Dance	11.8%	**88.2%**	100%
Sport	50%	50%	100%
TV	67%	33%	100%

Therefore, of those surveyed who enjoy Dance as their preferred leisure activity, 88.2% were women.

Segmented bar charts

The above comparison between the leisure activities and gender is even easier to make when presented in a visual manner. A **segmented bar chart** consists of one single bar for each **independent variable**, in this case gender. Each bar is then divided up into "**segments**", the size of which indicates the proportion of observations of the **dependent variable**, in this case leisure activity.



Identifying and analysing an association

Is there an **association** between the **leisure activity** and the **gender** in the previous example?

For there to be an association between the two categorical variables there has to be a **large or significant difference** between the percentage distribution of one set of categorical variable in response to the other categorical variable.

In the previous example, there is a significant difference in the chosen leisure activity depending upon the gender of people surveyed.

To fully answer the question "Is there an association between the leisure activity and the gender?" a **two-part response** is required:

Part 1 Identify the presence of the association
Yes, there is an association between the leisure activity chosen and gender

Part 2 Justify the presence of the association with appropriate percentages
The 10% of men who chose dance is much less than the 50% of women who chose dance.

OR

The 60% of men who chose sport is much greater than the 10% of women who chose sport

OR

The 30% of men who chose TV is much greater than the 40% of women who chose sport

Example 1

A survey of 150 peoples asked "If you could have a new vehicle, would you want a two door sports car?". Of the 100 men surveyed 80 wanted a two door sports car. Whereas, of the women surveyed, 40 didn't want the 2 door sports car.

Tasks

1. Construct a two way frequency table
2. Construct a two way percentage frequency table
3. Construct a segmented bar chart
4. Comment upon your findings

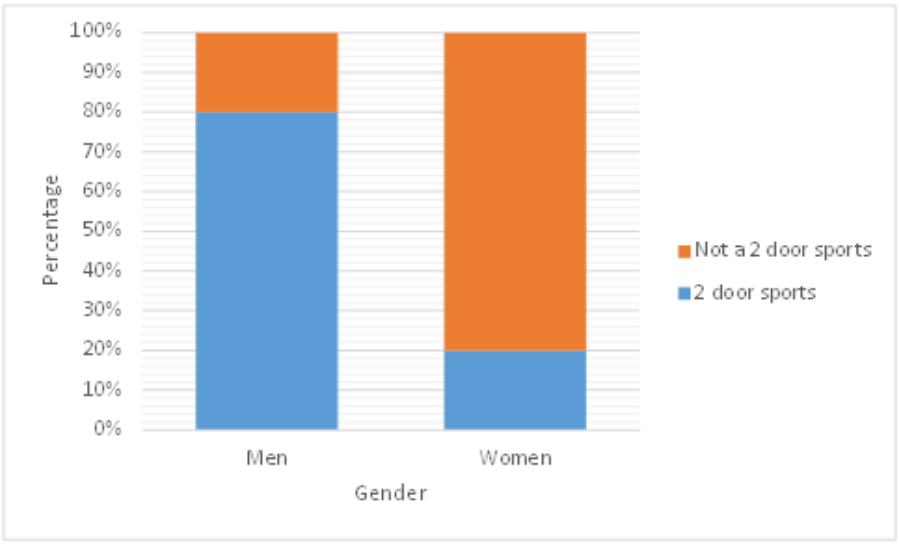
1. Two way frequency table

	Men	Women	Total
2 door sports	80	10	90
Not a 2 door sports	20	40	60
Total	100	50	150

2. Two way percentage frequency table

	Men	Women
2 door sports	80%	20%
Not a 2 door sports	20%	80%
Total	100%	100%

3. Segmented bar chart

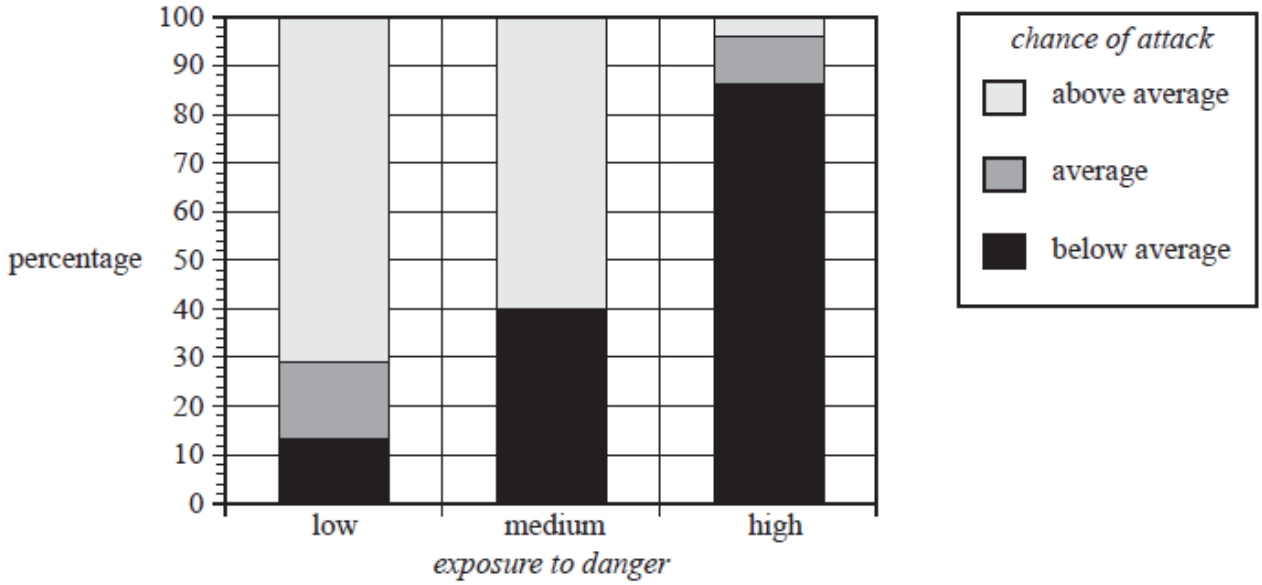


4. Comment

From the segmented bar chart it can be seen that 80% of the men surveyed wanted a two door sports car, whilst 80% of the women did not want a two door sports car.

Example 2

An animal study was conducted to investigate the relationship between *exposure to danger* during sleep (high, medium, low) and *chance of attack* (above average, average, below average). The results are summarised in the percentage segmented bar chart below.



The percentage of animals whose exposure to danger during sleep is high, and whose chance of attack is below average, is closest to

A. 4%
B. 12%
C. 28%
D. 72%
E. 86%

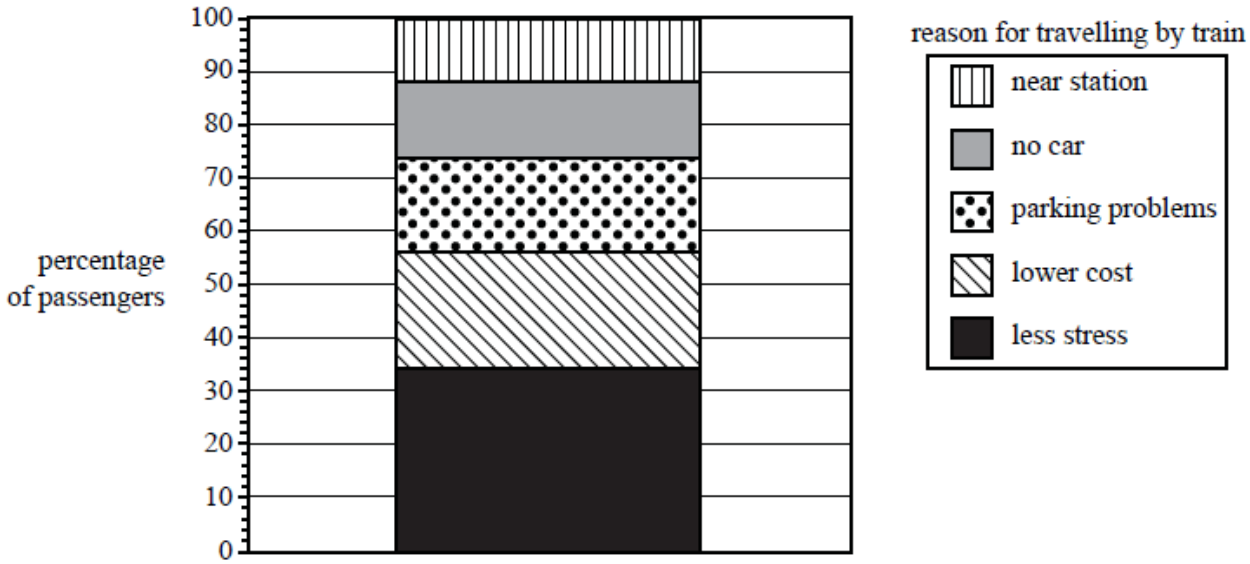
E

Animals whose exposure to danger is high is located in the right hand side percentage column. Those who have a below average chance of attack are coloured black. This represents a percentage of 80% +

∴ **Option E**

Example 3

The passengers on a train were asked why they travelled by train. Each reason, along with the percentage of passengers who gave that reason, is displayed in the segmented bar chart below.



The percentage of passengers who gave the reason 'no car' is closest to

A. 14%
B. 18%
C. 26%
D. 74%
E. 88%

A

Passengers who gave the reason "no car" are coloured grey in the above segmented bar chart. This represents the region from 74% to 88%. (88%-74% = 14%).

∴ Option A

Example 4

The table below shows the percentage of households with and without a computer at home for the years 2007, 2009 and 2011.

	Year		
	2007	2009	2011
Households with a computer	66.4%	77.7%	84.5%
Households without a computer	33.6%	22.3%	15.5%

In the year 2009, a total of 5 170 000 households were surveyed.
The number of households without a computer at home in 2009 was closest to

A. 801 000
B. 1 153 000
C. 1 737 000
D. 3 433 000
E. 4 017 000

B

From the table it can be seen that the percentage of households without a computer at home in 2009 is 22.3%.

$\therefore$ Actual numbers of households = $5\ 170\ 000 \times \frac{22.3}{100} = 1\ 152\ 910$

$\therefore$ **Option B**

Use the following information to answer Examples 5 and 6.

The heights of 82 mothers and their eldest daughters are classified as 'short', 'medium' or 'tall'. The results are displayed in the frequency table below.

		Mother		
		Short	Medium	Tall
Eldest daughter	Short	16	10	3
	Medium	8	14	11
	Tall	5	7	8

Example 5

The number of mothers whose height is classified as 'medium' is

A. 7
B. 10
C. 14
D. 31
E. 33

D

The total of the Mother "Medium" column is 10 + 14 + 7 = 31
∴ **Option D**

Example 6

Of the mothers whose height is classified as 'tall', the percentage who have eldest daughters whose height is classified as 'short' is closest to

A. 3%
B. 4%
C. 14%
D. 17%
E. 27%

C

The total of the Mother "Tall" column is 3 + 11 + 8 = 22
Of the 22 tall mother, 3 had a "Short" Eldest Daughter

∴ percentage of eldest daughter short from a tall mother = $\frac{3}{22} \times 100\%$ = 13.6% (Approx. 14%)
∴ **Option C**

Exam Styled Questions (current study design) – Multiple Choice

Use the following information to answer Questions 1 and 2.

The blood pressure (low, normal, high) and the age (under 50 years, 50 years or over) of 110 adults were recorded. The results are displayed in the two-way frequency table below.

Blood pressure	Age	
	Under 50 years	50 years or over
low	15	5
normal	32	24
high	11	23
Total	58	52

Question 1
(2016 Exam 1 Section A – Qn 1)
The percentage of adults under 50 years of age who have high blood pressure is closest to

A. 11%
B. 19%
C. 26%
D. 44%
E. 58%

B

The total of adults under 50 column is 15 + 32 +11 = 58
Of the 58 adults under 50, 11 have "high" blood pressure

$\therefore$ percentage of adult under 50 with high blood pressure = $\frac{11}{58} \times 100\%$ = 18.97% (Approx. 19%)
$\therefore$ **Option B**

Question 2
(2016 Exam 1 Section A – Qn 2)
The variables blood pressure (low, normal, high) and age (under 50 years, 50 years or over) are

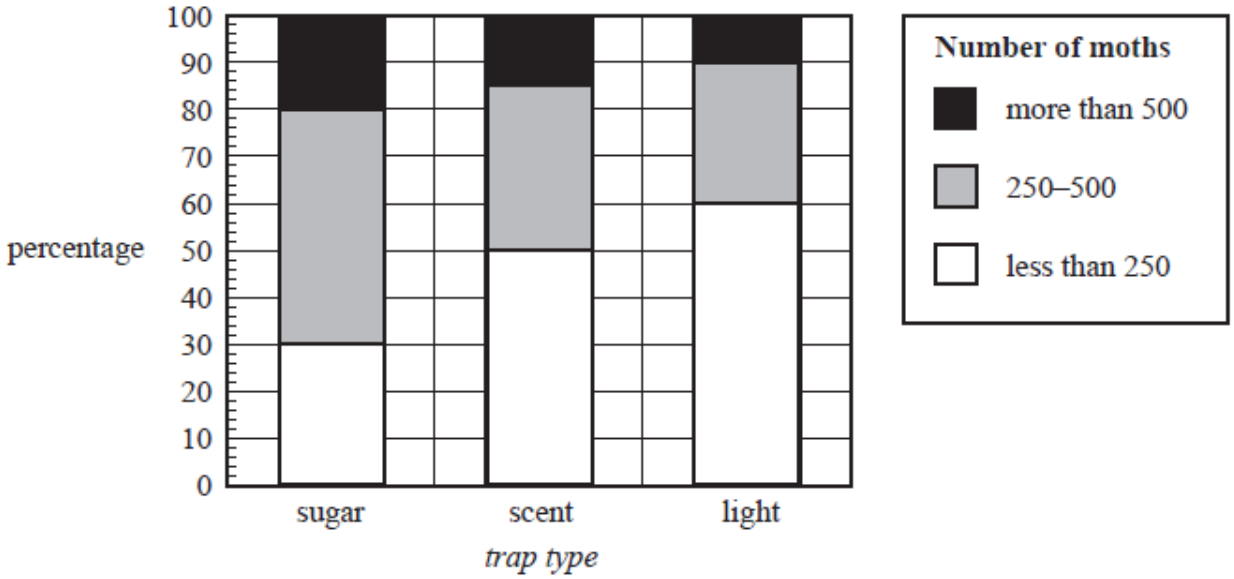
A. both nominal variables.
B. both ordinal variables.
C. a nominal variable and an ordinal variable respectively.
D. an ordinal variable and a nominal variable respectively.
E. a continuous variable and an ordinal variable respectively.

B

Both blood pressure and age are ordinal, they are ordered:
Blood pressure categories: low $\rightarrow$ normal $\rightarrow$ high (increasing order)
Age categories: Under 50 $\rightarrow$ 50 years and over (increasing order)
$\therefore$ **Option B**

Use the following information to answer Questions 3–5.

A study was conducted to investigate the association between the number of moths caught in a moth trap (less than 250, 250–500, more than 500) and the trap type (sugar, scent, light). The results are summarised in the percentaged segmented bar chart below.



Question 3

(2017 Exam 1 Section A – Qn 5)

There were 300 sugar traps.

The number of sugar traps that caught less than 250 moths is closest to

A. 30
B. 90
C. 250
D. 300
E. 500

B

The "sugar trap" bar is found on the far left bar. Number of moths less than 250 is coloured white. This represents 30% from off the graph.

$\therefore$ Actual numbers of sugar traps that caught less than 250 moths = $300 \times \frac{30}{100}$ = 90 moths

$\therefore$ **Option B**

Question 4
(2017 Exam 1 Section A – Qn 6)
The data displayed in the percentaged segmented bar chart supports the contention that there is an association between the number of moths caught in a moth trap and the trap type because

A. most of the light traps contained less than 250 moths.
B. 15% of the scent traps contained 500 or more moths.
C. the percentage of sugar traps containing more than 500 moths is greater than the percentage of scent traps containing less than 500 moths.
D. 20% of sugar traps contained more than 500 moths while 50% of light traps contained less than 250 moths.
E. 20% of sugar traps contained more than 500 moths while 10% of light traps contained more than 500 moths.

E

In order to answer this question we are looking for a "trend" between the two variables, that being the number of months caught and the type of trap.

Answer A is factually correct but doesn't compare the types of traps
Answer B is factually correct but doesn't compare the types of traps
Answer C is factually incorrect
Answer D is factually correct, but compares two different classification of number of moths across two different trap types.
Answer E is factually correct and compares a classification of number of moths across two different trap types.
∴ **Option E**

Question 5
(2017 Exam 1 Section A – Qn 7)
The variables number of moths (less than 250, 250–500, more than 500) and trap type (sugar, scent, light) are

A. both nominal variables.
B. both ordinal variables.
C. a numerical variable and a categorical variable respectively.
D. a nominal variable and an ordinal variable respectively.
E. an ordinal variable and a nominal variable respectively.

E

The variable "number of moths" is categorical. Whilst the variable "trap type" is nominal.
∴ **Option E**