gracesimpkins PERSONAL TUTORS

Suite 2, 179 Given Tce, Paddington, QLD, 4064
phone: 1300 360 245 | fax: 1300 360 237
email: contact@personaltutors.com.au
web: www.personaltutors.com.au

UNIT 3

Topic 1: Bivariate data analysis

1 Identifying and describing associations between two categorical variables

1.1 Define bivariate data.

Bivariate data – data showing the relationship between two variables.

But remember:

Data can be -

- Categorical data: data that are non-numerical but can be grouped, e.g. car types.
- Numerical data: data that can be either counted or measured, e.g. daily temperature.

Numerical data can be -

- Discrete data: can take only certain values such as whole numbers.
- Continuous data: can take any value within a certain range.

Data analysis can be -

- Univariate: study of only one variable at a time.
- Bivariate: study of two variables at the same time to determine if a relationship exists between those variables and how that relationship can be used to make predictions.

1.2 Construct two-way frequency tables and determine the associated row and column sums and percentages.

Two-way Frequency Tables for Bivariate Data Analysis

- Useful way of organising, identifying and describing an association between two categorical variables for a group of individuals or objects.
- Displays two variables which have been split into categories in a horizontal and a vertical direction.

gracesimpkins PERSONAL TUTORS

Suite 2, 179 Given Tce, Paddington, QLD, 4064
phone: 1300 360 245 | fax: 1300 360 237
email: contact@personaltutors.com.au
web: www.personaltutors.com.au

Using Given Values

- Calculate row and column totals.
- Answers will often be converted to percentages.
- It is important to know whether the row total, column total or overall total is what you are referencing.

E.g. What percentage of people with the disease tested positive?

- Use the disease row total as the denominator.

$$\frac{45}{50} \times 100 = 90\%$$

		Test Results		
		Positive	Negative	Total
Disease status	Disease	45	5	50
	No Disease	3	27	30
	Total	48	32	80

E.g. What percentage of people who tested positive have the disease?

- Use the positive column total as the denominator.

$$\frac{45}{48} \times 100 = 93.75\%$$

		Test Results		
		Positive	Negative	Total
Disease status	Disease	45	5	50
	No Disease	3	27	30
	Total	48	32	80

gracesimpkins PERSONAL TUTORS

Suite 2, 179 Given Tce, Paddington, QLD, 4064
phone: 1300 360 245 | fax: 1300 360 237
email: contact@personaltutors.com.au
web: www.personaltutors.com.au

Using Percentages

- Percentages will be based on one of the two categories.

E.g. based on disease status

- Use row totals.

		Test Results		
		Positive	Negative	Total
Disease status	Disease	$\frac{45}{50} \times 100 = 90\%$	$\frac{5}{50} \times 100 = 10\%$	100%
	No Disease	$\frac{3}{30} \times 100 = 10\%$	$\frac{27}{30} \times 100 = 90\%$	100%

E.g. based on test results

Use column totals.

		Test Results	
		Positive	Negative
Disease status	Disease	$\frac{45}{48} \times 100 = 93.75\%$	$\frac{5}{32} \times 100 = 15.625\%$
	No Disease	$\frac{3}{48} \times 100 = 6.25\%$	$\frac{27}{32} \times 100 = 84.375\%$
	Total	100%	100%

gracesimpkins PERSONAL TUTORS

Suite 2, 179 Given Tce, Paddington, QLD, 4064
phone: 1300 360 245 | fax: 1300 360 237
email: contact@personaltutors.com.au
web: www.personaltutors.com.au

1.3 Use an appropriately percentaged two-way frequency table to identify patterns that suggest the presence of an association.

- An association is the relationship between two variables.
- Percentaged two-way tables can show the association between variables.

Example – Is there an association between gender and full-time work participation?

	Full-time	Part-time	Total
Female	27%	73%	100%
Male	86%	14%	100%

Based on this sample, males are more likely to participate in full-time work. This is shown by 27% of females undertaking full-time work compared to 86% of males.

1.4 Understand an association in terms of differences observed in percentages across categories in a systematic and concise manner, and interpret this in the context of the data.

- For univariate data, if we know the value of one variable, we can generally predict another.
- For example, if half the people participating in training programs get a job, the likelihood of any one participant getting a job is 50%.
- If a second variable (bivariate data) is introduced, association can be determined from a percentage two-way table.

Example

Obtains a Job	**Length of Training Program**	
	Short	**Long**
No	100%	0%
Yes	0%	100%
Total	100%	100%

Suite 2, 179 Given Tce, Paddington, QLD, 4064
phone: 1300 360 245 | fax: 1300 360 237
email: contact@personaltutors.com.au
web: www.personaltutors.com.au

- If we know the length of the training program, we can perfectly predict the likelihood of getting a job.
- The longer the training program, the more likely the participant is to get a job and, conversely, the shorter the training program the less likely the participant is to get a job.
- That is, as the training program length increases, so does the likelihood of obtaining a job. The association is perfect.