

Exploring Data: Exploring Categorical Data—Frequency Tables

Descriptive statistics, including bivariate analysis, uses charts and tables to find relationships in sets of data. These relationships involve data that, for the most part, is quantitative. This chapter concentrates on data that is qualitative—**categorical data**. Examples of categorical data include religion, political party affiliation, gender, place of birth, citizenship status, and blood type. Quantitative data can be classified as categorical if ranges of values are grouped. For example, age is quantitative, but if ages are grouped, such as “Under 25” or “25 and Older,” then the data becomes categorical.

Marginal and Joint Frequencies for Two-Way Tables

Relationships between categorical variables are best shown using **contingency tables**. The relationships usually revolve around the dependence or independence of the variables.

EXAMPLE:

The gender breakdown for student enrollment at four colleges in a California community college district are as follows:

College A	Female = 3832	Male = 4228
College B	Female = 6765	Male = 5590
College C	Female = 2889	Male = 3388
College D	Female = 5580	Male = 5612

Organize this data into a two-way contingency table.

This table relates the two categorical variables, college and gender.

	<i>Female</i>	<i>Male</i>
College A	3832	4228
College B	6765	5590
College C	2889	3388
College D	5580	5612

The names of the colleges are called the row variable, and the gender is called the column variable.

Totals can be added to both the rows and columns. These totals are added in the margins so they are called marginal frequencies.

EXAMPLE:

Use the data from the previous example to create a two-way contingency table with marginal frequencies. Write the marginal frequencies as proportions and percentages of the totals:

	<i>Female</i>	<i>Male</i>	<i>TOTAL</i>
College A	3832	4228	8060
College B	6765	5590	12355
College C	2889	3388	6277
College D	5580	5612	11192
TOTAL	19066	18818	37884

Part II: Subject Area Reviews with Sample Questions and Answers

The marginal frequencies for the colleges are as follows:

$$\text{College A} = \frac{8060}{37884} = 21.3\%$$

$$\text{College B} = \frac{12355}{37884} = 32.6\%$$

$$\text{College C} = \frac{6277}{37884} = 16.6\%$$

$$\text{College D} = \frac{11192}{37884} = 29.5\%$$

The marginal frequencies for gender are as follows:

$$\text{Female} = \frac{19066}{37884} = 50.3\%$$

$$\text{Male} = \frac{18818}{37884} = 49.7\%$$

EXAMPLE:

Display the data in the previous example in bar graphs.

Bar graphs are used to display categorical data. Each graph displays the totals of one of the categories:

