-U

The distribution of 20 student by gender in an undergraduate CDA class in the spring 2020

Counts	Gender		Total	Nominal variable
	Male	Female		
n	08	12	20	

This table is reffered to as a one-way classification table bcz it was formed from the classification of a single variable.

Dichotomous / binary / Categorical - dichotomous

A variable that have only two categories is called Dichotomous or binary. And sometimes reffered to as the categorical-dichotomous

20 students in our CDA class classified by their economic status

Counts	Economic status			Total
	elite class	middle class	poverty level	
n_i	2	17	1	20

ordinal variable

The number of occurrences in each category is reffered to as the frequency count for that category

When subjects or objects are classified simultaneously by two or more attributes, the result of such a cross-classification can be conveniently arranged as a table of counts known as a contingency table

Cross-classification of students by gender & economic status:

Gender	Economic Status			Total
	Elite class	Middle class	Poverty class	
Male	1	6	1	08
Female	1	11	0	12
Total	2	17	01	20

2×3 table.

Here students have been cross-classified by two variables, gender and status. So the resulting table is known as two-way table or a two-way contigency table.
And if the students have been cross-classified by three variables then the resulting table will known as three way table or three way contigency table.

Cross-classification of students by gender, economic status and attitudes towards uni life.

2×3×3 table

Gender	Status	Attitudes			Total
		Positive	Mixed	negative	
Male	Elite	1	0	0	1
	Middle	3	1	2	6
	low	0	1	0	1
Femal	Elite	1	0	0	1
	middle	6	3	2	11
	low	0	0	0	0

In general, our data will consists of counts $\{n_i \Rightarrow i = 1, 2, \ldots, k\}$ in the K cells (of categories) for a contingency table.

These might be observations for the K levels of a single categorical variable or for $K = I \times J$ cells of a two-way $I \times J$ contingency table.

Categorical data consist of frequency counts of observations occurring in the response categories. Let X and Y denote two categorical variables, X having I levels and Y having J levels. We display the IJ possible combinations of outcomes in a rectangular table having I rows for the categories of X and J columns for the categories of Y. The cells of the table represent the IJ possible outcomes. A table of this form in which the cells contain frequency counts of outcomes is called a contingency table. A contingency table that cross classifies two variables is called a two-way table; one that cross classifies three variables is called a three-way table and so forth. A two way table having I rows and J columns an $I \times J$ (read I-by-J) table.

Probability structure for contingency tables:

Probabilies for contingency tables can be of three types - joint, marginal or conditional. (The sample proportions in the categories estimate the category probabilities)

Joint Probability:

A randomly chosen subject from the population of interest is classified on X and Y. Let $\pi_{ij} = P(X=i, Y=j)$ denote the probability that (X, Y) falls in the cell in row i and column j. The probabilities π_{ij} from the joint distribution of X and Y. That satisfy

$$\sum_i \sum_j \pi_{ij} = 1$$

$\pi_{ij} = n_{ij}/n$

Marginal Distribution.

The marginal distributions are the row and column totals of the joint probabilities. we denote these by π_{i+} for the row variable and π_{+j} for the column variable where the subscript "+" denotes the sum over the index it replaces. For 2x2 Tables.

$$\pi_{i+} = \pi_{11} + \pi_{12} \quad , \quad \pi_{+j} = \pi_{21} + \pi_{22}$$

Each marginal distribution refers to a single variable.

Conditional distribution:

In many contingency tables, one variable (say, the column variable, Y) is a response variable and the other (the row variable, X) is an explanatory variable. Then it is informative to construct a separate probability distribution of Y at each level of X. Such a distribution consists of conditional probabilities for Y, given the level of X. It is called a conditional distribution.

2x2 Table

Gender	Response Yes	No	Total
Male	$n_{11} = 02$	$n_{12} = 06$	08
Female	$n_{21} = 1$	$n_{22} = 11$	12
Total	03	17	20

Gender	Response Yes	No	$g(x)$
Male	$P_{11} = 0.1$	$P_{12} = 0.3$	$P_{1+} = 0.4$
Female	$P_{21} = 0.05$	$P_{22} = 0.55$	$P_{2+} = 0.6$
$h(y)$	$P_{+1} = 0.15$	$P_{+2} = 0.85$	1

For Male,

the proportion of "Yes" response

$$\frac{2}{8} = 0.25$$

the proportion of "No" response.

$$\frac{6}{8} = 0.75$$

The sample conditional distribution of having car is (0.25, 0.75)

For Female

The proportion of "Yes" response $= \frac{1}{12} = 0.08$

The proportion of "No" response $= \frac{11}{12} = 0.92$

The proportion (0.08, 0.92) from sample conditional distribution of having car.