



Part 1 –Marginal Distributions

At times, an individual can be categorized by two categorical variables. The data is displayed in a **two-way table**, where the independent variable is in the columns and the dependent variable is in the rows.

The data table below is an example of a two-way table. It is from a survey of New Yorkers from September 2008. This data can also be found on page 1.3 of the TI-Nspire document.

	Party Affiliation		
	Republican	Democrat	Independent
Obama	58	691	20
McCain	444	63	16
Other	5	8	1
Not Sure	21	23	3

When one variable is “ignored” and the proportion of the other variable that fits in each category is calculated, **marginal distributions** are found. There are two sets of marginal distributions for each two-way table.

Calculate the marginal distributions for the party affiliation variable.

Step 1: Find the sum of each column. The sum of Column B can be done two ways.

Method 1: In cell b5, type = **b1+b2+b3+b4**.

Method 2: In cell b5, type = **sum(b1:b4)**.

Step 2: Calculate the total number of respondents. To do this, title row 5 “total” and in cell e5, type a formula to find the sum of the three numbers in the row.

Step 3: Calculate the proportions of respondents for each affiliation. To do this, enter formulas in row 6 under each Column B–D. For the Republicans, the proportion formula is **=b5/e5**.

- What percent of the respondents are:

Republican: _____ Democratic: _____ Independent: _____

Calculate the marginal distributions for the candidates. Use column E and F. Write your answers in percents.

Obama: _____ McCain: _____

Other: _____ Not Sure: _____

- What can you infer from the marginal distribution for candidate choice?



Part 2 – Conditional Distributions

To study possible relationships between two categorical variables, one examines **conditional distributions**, the distributions of one variable for *given* categories of the other variable.

Restrict your attention to the Republicans. That is, only consider the respondents that declare themselves Republicans.

The conditional distribution for Obama is $\frac{58 \text{ (Obama \& Republican)}}{528 \text{ (Total Republicans)}} \approx .1098 \approx 11\%$.

Delete the marginal distributions from page 1.3. Calculate the conditional distribution given party affiliation.

- What percent of Republicans:
 favored Obama? _____ favored McCain? _____
 favored someone else? _____ were not sure? _____

Record the percentages in the chart below. Check your work by verifying that the column totals 100%.

Restrict your attention to the Democrats, then to the Independents, and fill in each space in the chart.

	Republican	Democrat	Independent
Obama			
McCain			
Other			
Not Sure			

- Pick a box from the table and explain what the percentage represents.

Now, restrict your attention to each candidate choice. Calculate the conditional distributions and fill in the chart below.

	Republican	Democrat	Independent
Obama			
McCain			
Other			
Not Sure			

- Pick a box from the table and explain what the percentage represents.



Homework

The following table of data shows responses to a poll asking the question "In general, do you favor or oppose the building of more nuclear power plants in the United States?"

	Echo Boomers (18–31)	Gen X (32–43)	Baby Boomers (42–62)	Matures (63+)
Strongly favor	48	62	55	57
Somewhat favor	60	94	61	47
Somewhat oppose	64	57	33	24
Strongly oppose	51	59	35	16
Not sure	81	63	35	19

Calculate the marginal distributions for age and opinion. Then, calculate the conditional distributions, given age and opinion. Create a chart similar to the one in the problem above.