MATH 143: Introduction to Probability and Statistics

Worksheet for Fri., Sept. 10: Two-Way Tables

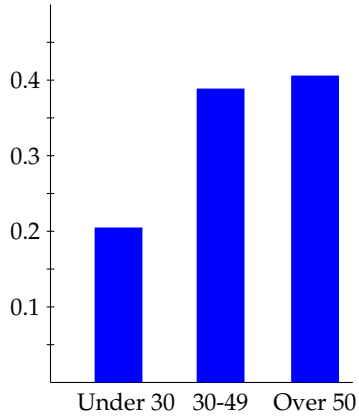
In a survey of adult Americans in 1986, each respondent was asked to indicate his/her age and to categorize his/her political ideology. The results are summarized in the following table of counts:

	Political Ideology			
Age Range	Liberal	Moderate	Conservative	Total
Under 30	83	140	73	296
30–49	119	280	161	560
Over 50	88	284	214	586
Total	290	704	448	1442

This table is called a *two-way table* (or *contingency*) since it classifies each person according to two categorical variables. In particular, it is a 3×3 table; the first number represents the number of categories of the row variable (age), and the second number represents the number of categories of the column variable (political ideology).

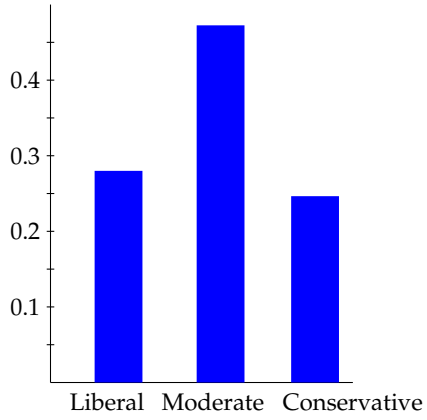
1. How many survey respondents in the "Over 50" age group consider themselves politically "moderate"? 284

2. What *proportion* of respondents were under the age of 30? 296/1442 ≐ 0.205 What proportion between 30 and 50? 560/1442 ≐ 0.388 Over 50? 586/1442 ≐ 0.406 Use the grid at right to draw a bar graph conveying this information. You have found the distribution for the *age* variable. In relation to this two-way table, this distribution is called a *marginal distribution*, as the important values appear "in the margin" (ignoring any breakdown by political ideology).



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

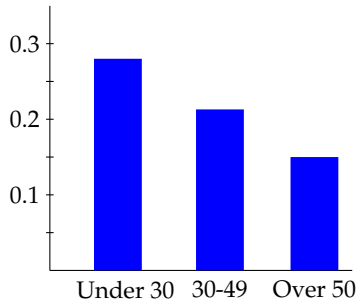
3. Restrict your attention for the moment to respondents under 30 years of age. What proportion of these young respondents classify themselves as *liberal*? 83/296 ≐ 0.28 As moderate? 0.473 As conservative? 0.247 Plot these proportions as a bar graph at right. You have found the *conditional distribution of political ideology for respondents under the age of 30*.



4. Fill in the top row of the table below with the proportions you found in the previous problem; once again, this row gives the distribution (among respondents) for political ideology in the *under 30* age group. There are two more conditional distributions for political ideology given a fixed age group. Find them, and fill in the remaining two rows.

	Proportion of		
Age	Liberals	Moderates	Conservatives
Under 30	0.28	0.473	0.247
30–49	0.213	0.5	0.288
Over 50	0.15	0.485	0.365

5. Use the space at right to draw another bar graph. Make one bar for each age group, with the height of the bar representing the proportion of liberals within that age group. Is this a conditional distribution? No (Do the heights of the bars add up to 1?)



6. Compare proportions of liberals across age groups. Do the same for moderates and conservatives. Does there seem to be a relationship between age and political ideology? It seems there is a general movement towards conservativism as poeple age (so the variables are related). More specifically, some liberals move toward being moderate, while near equal numbers of moderates have become conservatives.

Sometimes there is good reason to take data from a two-way table and break it down further by introducing another categorical variable. For instance, a paper from 1981 investigating racial biases in the application of the death penalty investigated 326 cases in which the defendant was convicted of murder. Here is a breakdown of these cases by *race of defendant* and *whether or not the defendant was sentenced to death*.

	Death Penalty?	
Race	Yes	No
Black	17	149
White	19	141

But when we look at another variable, the *race of the victim*, we see a different story.

Black victim

	Death Penalty?	
Race	Yes	No
Black	6	97
White	0	9

White victim

	Death Penalty?	
Race	Yes	No
Black	11	52
White	19	132

This type of *reversal* phenomenon is known as **Simpson's paradox**.