Continuing coverage of inferences for yes/no data using sample and population proportions.

2.5. Relations in categorical data

1. **two-way tables** describe two categorical variables using
   a. frequency counts or proportions (percentages)
      (1) rounding error may make columns or rows not quite add
   b. **row and column variables**

2. **marginal distributions** consider one variable (rows or columns, only) at a time
   a. bar graphs are natural modes of display
   b. distributions are more effective when frequencies are converted to proportions

3. describing relationships
   a. in general “relationship” or “association”, means that change in the value of one
      variables corresponds to a change in the value of another variable
      (1) there is not necessarily a causal relationship
   b. kinds of relationships between events. Examples include …
      (1) disjoint: pairs of table cells are disjoint relative to one another
      (2) joint: each table cell frequency relative to the entire sample frequency
      (3) conditional: each table cell frequency relative to its row or column sample frequency
   c. normally we focus on conditional relations via **conditional distributions**, to
      investigate whether a relationship “association” exists between the row and column
      variable

4. **conditional distributions**
   a. together,
      (1) the conditional column proportions comprise a conditional distribution for their
         respective column value, and
      (2) the conditional row proportions comprise a conditional distribution for their
         respective row value
   b. if there are \(r\) rows and \(c\) columns, then there are …
      (1) \(r\) conditional distributions for the \(c\) column variable values, and
      (2) \(c\) conditional distributions for the \(r\) row variable values
   c. to detect an association we compare either the corresponding proportions in the …
      (1) \(r\) conditional distributions for the \(c\) column variable values, or
      (2) \(c\) conditional distributions for the \(r\) row variable values
   d. if the compared proportions are different for a particular variable value, this indicates an
      “association” between the values of the other variable in the context of to the particular
      variable value
      any such association is said to indicate an association between the row and column
      variable, not just the particular variable value and the other variable

5. Simpson’s paradox
   a. lurking variables can change or even reverse the relationship between variables
      (1) this is true for the relationship between categorical variables, as well as, quantitative
          variables
      (2) in the case of categorical variables this phenomenon was popularized by a man
          named Simpson
   b. textbook examples demonstrate the possibility of a third variable altering the apparent
      relationship between two categorical variables
What proportion of survey respondents are under age 30?
What proportion of survey respondents are between 30 and 50 years of age?
What proportion of survey respondents are over age 50?

### Frequencies

<table>
<thead>
<tr>
<th></th>
<th>Under 30</th>
<th>30-49</th>
<th>Over 50</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liberal</td>
<td>83</td>
<td>119</td>
<td>88</td>
<td>290</td>
</tr>
<tr>
<td>Moderate</td>
<td>140</td>
<td>280</td>
<td>284</td>
<td>704</td>
</tr>
<tr>
<td>Conservative</td>
<td>73</td>
<td>161</td>
<td>214</td>
<td>448</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>296</td>
<td>560</td>
<td>586</td>
<td>1442</td>
</tr>
</tbody>
</table>

What proportion of survey respondents are under age 30 and liberal?
What proportion of survey respondents are between 30 and 50 years of age and moderate?
What proportion of survey respondents are over age 50 and conservative?

### Joint distribution

<table>
<thead>
<tr>
<th></th>
<th>Under 30</th>
<th>30-49</th>
<th>Over 50</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liberal</td>
<td>83/1442</td>
<td>119/1442</td>
<td>88/1442</td>
<td>290</td>
</tr>
<tr>
<td>Moderate</td>
<td>140/1442</td>
<td>280/1442</td>
<td>284/1442</td>
<td>704</td>
</tr>
<tr>
<td>Conservative</td>
<td>73/1442</td>
<td>161/1442</td>
<td>214/1442</td>
<td>448</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>296</td>
<td>560</td>
<td>586</td>
<td>1442</td>
</tr>
</tbody>
</table>
What proportion of young respondents classify themselves as liberal?
What proportion of young respondents classify themselves as moderate?
What proportion of young respondents classify themselves as conservative?

What proportion of respondents between 30 and 50 years of age classify themselves as conservative?
What proportion of older respondents classify themselves as conservative?

### Conditional distributions of political ideology for three age groups

<table>
<thead>
<tr>
<th></th>
<th>Under 30</th>
<th>30-49</th>
<th>Over 50</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liberal</td>
<td>83/296</td>
<td>119/560</td>
<td>88/586</td>
<td>290</td>
</tr>
<tr>
<td>Moderate</td>
<td>140/296</td>
<td>280/560</td>
<td>284/586</td>
<td>704</td>
</tr>
<tr>
<td>Conservative</td>
<td>73/296</td>
<td>161/560</td>
<td>214/586</td>
<td>448</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>296</td>
<td>560</td>
<td>586</td>
<td>1442</td>
</tr>
</tbody>
</table>

### Conditional distributions of political ideology for three age groups

<table>
<thead>
<tr>
<th></th>
<th>Under 30</th>
<th>30-49</th>
<th>Over 50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liberal</td>
<td>28.0%</td>
<td>21.3%</td>
<td>15.0%</td>
</tr>
<tr>
<td>Moderate</td>
<td>47.3%</td>
<td>50.0%</td>
<td>48.5%</td>
</tr>
<tr>
<td>Conservative</td>
<td>24.7%</td>
<td>28.8%</td>
<td>36.5%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

What proportion of respondents who classify themselves as liberal, are young?