Two-way arrays.

- two-way table
- two-way layout
- two-factor array
- contingency tables

New data type/structure (for course)

Rectangular display

- rows, columns, responses

- different factors/classifications vary separately

- response for each combination of levels of the factors

\[ Y_{ij} \quad i=1,\ldots,I; \quad j=1,\ldots,J \]

\[ n = I \times J \text{ observations (cells)} \]

\[ y \text{ numerical} \]
Factors may be labels, ordered, numerical

Interested in relation between response and rows and columns
   wish summary highlighting relation between response and each factor

Example - area burned in wildfires by month and year

Question - prediction?

the data

row is month, column is year (92-02)
   I = 12, J = 10

(months have differing numbers of days)

boxplots for rows, columns

Conceptualization.

Response

   \approx \text{summary} + \text{row effect} + \text{column effect}
\[ y_{ij} \approx \mu + \alpha_i + \beta_j \]

Separate contribution for each factor

Additive dependence

(May need to transform. Later)

Old \( \beta \) is now \( \theta = (\mu, \alpha_i, \beta_j) \)

Paradigm.

\[ \text{data} = \text{fit} + \text{residual} \]

Fitting.

OLS

\[ \min_{\theta} \sum_{i,j} (y_{ij} - \mu - \alpha_i - \beta_j)^2 \]

overparametrized

side conditions

\[ \sum_i \alpha_i, \sum_j \beta_j = 0 \]
normal equations

\[ m = \bar{y}, \quad a_i = (\bar{y}_i. - \bar{y}), \quad b_j = (\bar{y}_j - \bar{y}) \]

ANOVA identity

\[ \sum_i \sum_j y_{ij}^2 \]

\[ = \sum_i \sum_j (\bar{y})^2 + \sum_i \sum_j (\bar{y}_i. - \bar{y})^2 + \sum_i \sum_j (\bar{y}_j - \bar{y})^2 \]

from orthogonality relations

ANOVA TABLE

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>( \sum_i \sum_j (\bar{y})^2 )</td>
<td>1</td>
</tr>
<tr>
<td>rows</td>
<td>( \sum_i \sum_j (\bar{y}_i. - \bar{y})^2 )</td>
<td>(I-1)</td>
</tr>
<tr>
<td>columns</td>
<td>( \sum_i \sum_j (\bar{y}_j - \bar{y})^2 )</td>
<td>(J-1)</td>
</tr>
<tr>
<td>residual</td>
<td>( \sum_i \sum_j (y_{ij} - \bar{y}_i. - \bar{y}_j + \bar{y})^2 )</td>
<td>(I-1)(J-1)</td>
</tr>
</tbody>
</table>
total \ \sum_i \sum_j y_{ij}^2 \ \ \ \text{n = IJ}

Wildfire example.

Plot effects \(a_i, b_j\) (parallel boxplots)

**ANOVA TABLE**

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>14.912</td>
<td>1</td>
</tr>
<tr>
<td>rows</td>
<td>17.270</td>
<td>11</td>
</tr>
<tr>
<td>columns</td>
<td>3.720</td>
<td>9</td>
</tr>
<tr>
<td>residual</td>
<td>10.642</td>
<td>99</td>
</tr>
<tr>
<td>total</td>
<td>46.544</td>
<td>120</td>
</tr>
</tbody>
</table>

twoway(trim=0), aov()

Response may be summary of a batch

Finding patterns difficult with large tables
If classical test rejects, what next? EDA can suggest