One-Factor Experiments

Assume we want to investigate the effects of a single factor, with multiple levels and \( r \) observations at each such level.

\[
y_{ij} = \mu + \alpha_j + e_{ij}
\]

where \( \mu \) is mean response, \( \alpha_j \) is the effect of alternative \( j \) and \( e_{ij} \) is the error term.

We choose to scale things such that the \( \alpha_j \)'s sum to zero.

\[
\begin{array}{ccc}
144 & 101 & 130 \\
120 & 144 & 180 \\
176 & 211 & 141 \\
288 & 288 & 374 \\
144 & 72 & 302 & \text{Sum} \\
\hline
Col Sum & 872 & 816 & 1127 & 2815 \\
Col Mean & 174.4 & 163.2 & 225.4 & 187.7 \\
Effect & -13.3 & -24.5 & 37.7 & 0
\end{array}
\]
Errors and Variation

For each observation, the error is the difference between the observation and the sum of the mean and alternative effect.

- Mean is 187.7, R’s effect is -13.3, so the first error term is
  \[ 144 - (187.7 - 13.3) = -30.4. \]

If we sum the squares of all the error terms, we get 94,365.20.

If we sum the squares of all the responses, we get 105,357.3

As usual, the total variation from the mean can be allocated to the effect of the factor and the effect of the error.

- \[ \text{SST} = \text{SSY} - \text{SS0} = \text{SSA} + \text{SSE} \]
- Variation is divided into explained and unexplained error

The error accounts for

\[ \frac{94,365.20}{105,357.3} = 89.6\% \text{ of the variation.} \]

- The factor accounts for 10.4\% of the variation.
- Is this significant?
ANOVA

We use ANOVA to determine if there is a meaningful difference due to this factor.

Recall that we have to scale the variations by the degrees of freedom.

\[
SSY = SS0 + SSA + SSE
\]
\[
ar = 1 + (a - 1) + a(r - 1)
\]

- We have \( ar \) independent terms (for \( SSY \))
- One mean (for \( SS0 \))
- \( a \) experiments but only \( a - 1 \) independent ones since they add to zero
- \( ar \) error terms, but since the \( r \) errors for a given experiment add to zero, each of \( a \) experiments has only \( r - 1 \) degrees of freedom.

Calculate the ratio of \( SSA/DOF(a) \) and \( SSE/DOF(e) \)

- \((10,992.13/2)/(94,365.2/12) = 0.7\)
- From the F table position \([2, 12]\) we require a value of 2.81

Conclusion: This factor does not give a significant difference
Two-Factor Full Factorial Design

<table>
<thead>
<tr>
<th>Work</th>
<th>1</th>
<th>2</th>
<th>No</th>
<th>Sum</th>
<th>Mean</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASM</td>
<td>54</td>
<td>55</td>
<td>106</td>
<td>215</td>
<td>71.7</td>
<td>-0.5</td>
</tr>
<tr>
<td>TECO</td>
<td>60</td>
<td>60</td>
<td>123</td>
<td>243</td>
<td>81.0</td>
<td>8.8</td>
</tr>
<tr>
<td>SIEVE</td>
<td>43</td>
<td>43</td>
<td>120</td>
<td>206</td>
<td>68.7</td>
<td>-3.5</td>
</tr>
<tr>
<td>DHRY</td>
<td>49</td>
<td>52</td>
<td>111</td>
<td>212</td>
<td>70.7</td>
<td>-1.5</td>
</tr>
<tr>
<td>SORT</td>
<td>49</td>
<td>50</td>
<td>108</td>
<td>207</td>
<td>69.0</td>
<td>-3.2</td>
</tr>
<tr>
<td>Sum</td>
<td>255</td>
<td>260</td>
<td>568</td>
<td>1083</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>51.0</td>
<td>52.0</td>
<td>113.6</td>
<td></td>
<td>72.2</td>
<td></td>
</tr>
<tr>
<td>Effect</td>
<td>-21.2</td>
<td>-20.2</td>
<td>41.4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We can now compute error terms using the model

\[ y_{ij} = \mu + \alpha_j + \beta_i + e_{ij} \]

Variances:
- SSA is the sum of squares of column effects = 12,857.2
- SSB is the sum of squares of row effects = 308.40
- SSE can be computed from the table of errors, or as SST – SSA – SSB, = 236.80

Caches (A) explain 95.9% of variation, workloads (B) explain 2.3%, and errors explain 1.8%
ANOVA

DOF

- For caches, $3 - 1 = 2$, so
  $12,857.2/2 = 6428.6$
- For workloads, $5 - 1 = 4$ so $308.4/4 = 77.10$
- For errors, $2 \times 4 = 8$ so $236.80/8 = 29.60$

F-Table

- $6428.6/29.60 = 217.2$, higher than the required 3.1
- $77.10/29.60 = 2.6$, lower than the required 2.8
- This is good, workloads are not meant to be a confounding influence
Precautions

If there are any doubts, then:

- Check that the errors are normal (use quantile-quantile plot)
- Check confidence intervals on the parameters, especially mean and on effects of the supposed significant factors.

In this case, the quantile-quantile plot (Figure 21.1) looks reasonable.

In this case, we get confidence intervals of:

- Grand mean: 72.2 ± 2.6
- Cache means: All effects do not include zero in the confidence interval
- Workloads: Only TECO does not include zero in the confidence interval, indicating that they are indistinguishable
- This is good: Workload was meant not to confound the cache study
Case Study

Problem: What factors affect the pedagogical effectiveness of algorithm visualizations?

Purvi Saraiya, Cliff Shaffer, Scott McCrickard, Chris North

Context:

- There are 10 Gazillion java applets for algorithm/data structure visualization
- There are 1 Gazillion papers about algorithm visualizations
- 10% of them did enough analysis to claim that the students “like” the visualization, but can determine no significant effect over lecture or book
- .1% are good enough and have conducted analysis to claim a significant effect
Identifying Factors

Used “Expert review panel” (ourselves) to evaluate a collection of heap sort visualizations

- Goal: Identify key factors, (expert judgement)
- Several visualizations of one algorithm
- Noted both good and bad points to each visualization
- Synthesized observations into feature list

Features

- High usability, no bugs
- Student data input
- Appropriate feedback messages (background appropriate)
- User control (vs. animation)
- Backing up
- State changes are clear
- Multiple views (physical vs. logical views)
- Window management/relationships
- Pseudocode
- Guided (questions to answer)

Some of these were determined to be “baseline” requirements, not to be directly studied for their significance.
Features of Interest

- Student input
- Sample given
- User control vs. animation
- Pseudocode
- Back button
- Guide