Understanding Factors Influencing the Citation Count for Conference Papers

Utkarsh Goel
Sandhya Thangavel
Motivation
Null Hypothesis
Null Hypothesis
Alternate Hypothesis
Response Variable and Factors

● Response variable
  ○ Citation count

● Factors
  ○ Choice of keywords in the paper title
    ■ Cloud, Distributed, Security, Mobile, Performance
  ○ Conference Rank
    ■ A, B, C
  ○ Publication year
    ■ 2008 until 2012 (5 years)
Research Questions

• How to **classify conferences** in different ranks?

• Is there an **accurate database** for conference rankings?

• Does the rankings **change over time**?
Data Collection Technique

- Google Survey for determining conference ranking
  Insight of Community opinion (47 responses)

- Online Conference Ranking Databases
  Core Conference Portal, ERA Ranking Portal

- Prof. Kevin Almeroth’s blog on Conference Statistics
  Acceptance rate, #Papers submitted, #Papers accepted

- Google Scholar
  Paper Title, Citation Count, Conference Name, Publication Year
Data Collected

• 63 networking conferences
  22 rank A conferences (Acceptance Rate: 11% to 20.9%)
  22 rank B conferences (Acceptance Rate: 21% to 30.9%)
  19 rank C conferences (Acceptance Rate: 31% to 40.9%)

• Total papers = 38895
  #Papers in Rank A Conferences = 8224
  #Papers in Rank B Conferences = 7134
  #Papers in Rank C Conferences = 23537
Research Questions

• How to classify conferences in different ranks?

• Is there an **accurate database** for conference rankings?

• Does the rankings change over time?
• What
Variation in Conference Ranking
Research Questions

• How to classify conferences in different ranks?
• Is there an accurate database for conference rankings?
• Does conference rankings change over time?
Conference Rankings over Time

Percentage of Conferences (%)

- Rank A
- Rank B
- Rank C


Mountains & Minds
How do people cite papers?

- Search papers from high quality conferences?
- Ask Google?
- Search papers from an year?
Does Conference affect Citation Count?
Does Paper Title affect Citation Count?

- Cloud
- Distributed
- Security
- Mobile
- Performance

Average Citations
Does Publication Year affect Citation Count?

![Box plot showing average citations per year from 2008 to 2012.](image)
# 3-Way Nested ANOVA

<table>
<thead>
<tr>
<th>Keyword/Year</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloud</td>
<td>494.2</td>
<td>297.2</td>
<td>155.5</td>
</tr>
<tr>
<td>Performance</td>
<td>57.3</td>
<td>43.3</td>
<td>50.9</td>
</tr>
<tr>
<td>Distributed</td>
<td>31</td>
<td>29.3</td>
<td>39.7</td>
</tr>
<tr>
<td>Security</td>
<td>42.8</td>
<td>61.6</td>
<td>67.3</td>
</tr>
<tr>
<td>Mobile</td>
<td>67.4</td>
<td>75</td>
<td>46.9</td>
</tr>
</tbody>
</table>
3-Way Nested ANOVA Contd...

```r
> aov.out = aov(myDF$citations ~ myDF$rank + myDF$year + myDF$title + myDF$rank * myDF$year + myDF$rank * myDF$title + myDF$year * myDF$title + myDF$rank * myDF$year * myDF$title, data = myDF)
> summary(aov.out)

             Df Sum Sq Mean Sq  F value Pr(>F)
myDF$rank    1  44533  44533 19.137   4.35e-05  ***
myDF$year    1  30365  30365 13.049   0.000581  ***
myDF$title   1  31694  31694 13.620   0.000451  ***
myDF$rank:myDF$year  1  17546  17546  7.540   0.007742  **
myDF$rank:myDF$title  1  17276  17276  7.424   0.008202  **
myDF$year:myDF$title  1  26315  26315 11.308   0.001279  **
myDF$rank:myDF$year:myDF$title  1  17031  17031  7.319   0.008645  **
Residuals    67 155915  2327
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```
Conclusion

- **Conference Ranking** affects citation count by 13%.
- **Paper title** affects citation count by 9.3%.
- **Publication year** affects citation count by 9%.
- **Interactions** affect citation count by 23%.
- **Unknown factors** affect citation count by 45.7%.
Conclusion Contd…

• What else could affect citations?
Conclusion Contd…

• What else could affect citations?
  • Author popularity.
Conclusion Contd...

• What else could affect citations?
  • Author popularity.
  • Paper relevance.
Conclusion Contd...

- What else could affect citations?
  - Author popularity.
  - Paper relevance.
  - Timeliness of paper.
Conclusion Contd…

• What else could affect citations?
  • Author popularity.
  • Paper relevance.
  • Timeliness of paper.
  • Citation source.
Conclusion Contd...

- What else could affect citations?
  - Author popularity.
  - Paper relevance.
  - Timeliness of paper.
  - Citation source.
  - Access to papers.

"We didn't read half of the papers we cite because they are behind a paywall" 
#overlyhonestmethods
Conclusion Contd…

Data and Code Available at:
https://github.com/ugoel/ESOF522