Attendance estimation for massive demonstrations using basic statistics.

Annual meeting of the SEIO Working Group in Multivariate Analysis and Classification (AMyC)

Llorenç Badiella

1Servei d’Estadística Aplicada
Universitat Autònoma de Barcelona
E-mail: llorenc.badiella@uab.cat

Barcelona, October 20-21, 2016.
Whenever there is a demonstration, the same dilemma arises:

**What was the number of participants?**

**Preliminaries**

- The number of participants is used as an indicator of the success of the event.
- Data provided by the organization usually has a credibility near to zero.
- When politics are involved, the figures from "official sources" show alarming discrepancies, according to their affinity with the call.
- Official data show severe inconsistencies between concentrations.
- The effort of the event is exposed to skepticism.
The problem

Q: How many elephants will fit into a Mini?
A: Four: Two in the front, two in the back.
Q: How many elephants will fit into a Mini?
A: Four: Two in the front, two in the back.
Can we obtain accurate, valid and reproducible estimates?
1977: Llibertat, Amnistia, Estatut d’Autonomia

La Vanguardia

<table>
<thead>
<tr>
<th>Source</th>
<th>Estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organization</td>
<td>1.000.000</td>
</tr>
<tr>
<td>Official</td>
<td>1.000.000</td>
</tr>
</tbody>
</table>

Attendance estimation for massive demonstrations
1977: Llibertat, Amnistia, Estatut d’Autonomia

<table>
<thead>
<tr>
<th>Source</th>
<th>Estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organization</td>
<td>1.000.000</td>
</tr>
<tr>
<td>Official</td>
<td>1.000.000</td>
</tr>
</tbody>
</table>
2003: Aturem la Guerra

tercerainformacion.es

<table>
<thead>
<tr>
<th>Source</th>
<th>Estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organization</td>
<td>1,500,000</td>
</tr>
<tr>
<td>Guàrdia Urbana</td>
<td>1,300,000</td>
</tr>
<tr>
<td>Deleg. Gobierno</td>
<td>350,000</td>
</tr>
</tbody>
</table>

L. Badiella  

Attendance estimation for massive demonstrations
### 2003: Aturem la Guerra

**Source** | **Estimation**
---|---
Organization | 1.500.000
Guàrida Urbana | 1.300.000
Deleg. Gobierno | 350.000

tercerainformacion.es

L. Badiella  
Attendance estimation for massive demonstrations
Attendance estimation for massive demonstrations

<table>
<thead>
<tr>
<th>Source</th>
<th>Estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organization</td>
<td>1,500,000</td>
</tr>
<tr>
<td>Guàrdia Urbana</td>
<td>1,100,000</td>
</tr>
<tr>
<td>El País</td>
<td>425,000</td>
</tr>
<tr>
<td>Agència EFE (Lynce)</td>
<td>62,000</td>
</tr>
</tbody>
</table>

2010: Sentència del TC sobre l’Estatut

Wikipedia
### Attendance estimation for massive demonstrations

<table>
<thead>
<tr>
<th>Source</th>
<th>Estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organization</td>
<td>1,500,000</td>
</tr>
<tr>
<td>Guàrida Urbana</td>
<td>1,100,000</td>
</tr>
<tr>
<td>El País</td>
<td>425,000</td>
</tr>
<tr>
<td>Agència EFE (Lynce)</td>
<td>62,000</td>
</tr>
</tbody>
</table>
2012: Catalunya nou Estat d’Europa

La Vanguardia

<table>
<thead>
<tr>
<th>Source</th>
<th>Estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organization</td>
<td>2,000,000</td>
</tr>
<tr>
<td>Guàrida Urbana</td>
<td>1,500,000</td>
</tr>
<tr>
<td>Deptm. Interior</td>
<td>1,500,000</td>
</tr>
<tr>
<td>Deleg. Gobierno</td>
<td>600,000</td>
</tr>
</tbody>
</table>
2012: Catalunya nou Estat d’Europa

La Vanguardia

<table>
<thead>
<tr>
<th>Source</th>
<th>Estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organization</td>
<td>2.000.000</td>
</tr>
<tr>
<td>Guàrida Urbana</td>
<td>1.500.000</td>
</tr>
<tr>
<td>Deptm. Interior</td>
<td>1.500.000</td>
</tr>
<tr>
<td>Deleg. Gobierno</td>
<td>600.000</td>
</tr>
</tbody>
</table>
### 2013: Día Hispanidad

Font: La Vanguardia

<table>
<thead>
<tr>
<th>Source</th>
<th>Estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organization</td>
<td>160,000</td>
</tr>
<tr>
<td>Deleg. Gobierno</td>
<td>105,000</td>
</tr>
<tr>
<td>Deptm. Interior</td>
<td>30,000</td>
</tr>
</tbody>
</table>

Attendance estimation for massive demonstrations
### Source Estimation Organization

<table>
<thead>
<tr>
<th>Source</th>
<th>Estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organization</td>
<td>160,000</td>
</tr>
<tr>
<td>Deleg. Gobierno</td>
<td>105,000</td>
</tr>
<tr>
<td>Deptm. Interior</td>
<td>30,000</td>
</tr>
</tbody>
</table>
2007: España por la libertad. No más cesiones a ETA

<table>
<thead>
<tr>
<th>Source</th>
<th>Estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Com. Madrid</td>
<td>2.125.000</td>
</tr>
<tr>
<td>Del. Gobierno</td>
<td>342.655</td>
</tr>
</tbody>
</table>
2007: España por la libertad. No más cesiones a ETA

<table>
<thead>
<tr>
<th>Source</th>
<th>Estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Com. Madrid</td>
<td>2.125.000</td>
</tr>
<tr>
<td>Del. Gobierno</td>
<td>342.655</td>
</tr>
</tbody>
</table>

Cadenaser.com

L. Badiella

Attendance estimation for massive demonstrations
Classical Method

Jacobs Method (1967)

- Area × Density.


New Formula Counts Crowds

The quadrennial presidential election campaigns are not only a time for inflated speeches. They also herald a return of that phenomenon known as inflated crowd estimating — or deflated estimating, depending upon which party you listen to.

Now a new development may throw some honesty into the latter preoccupation of politicians.

Herbert Jacobs, a lecturer in journalism at the University of California, has devised what he calls “Jacobs’ crowd formula.” The formula calls for walking off the length and width of a gathering, then adding the two figures. If the crowd is densely packed, this is multiplied by 10; if loosely assembled, by seven.

The average person occupies from six to eight square feet in most crowds, he says. For sit-ins, the figure is 5.7 square feet.

Fable Of Night Births

Cold statistics have abolished another myth. The latest is the one that holds that babies are almost always born in the wee, small hours of the morning.

"Tain't so. Women's Medical News Service reports on a study of births in New York City by doctor of science Carl L. Erhard.

He found that only 35 out of 1,362 babies were born between 2 and 3 a.m. The largest number were born between noon and 1 p.m.

For some reason, first babies are likely to be born at any time of day or night. But if the mother has had previous children, subsequent little stranglers are more likely to arrive between 9 a.m. and noon.

"By no means are most babies born in the middle of the night," says Erhardt. "Mothers can now relax."

Cormley In Washington

A Lesson In Appalachia

By Ray Crowley

WASHINGTON (NEA) — A strange thing happened to two psychiatrists who went with a team to help “unfortunate, deprived” children in poverty-stricken areas of Appalachia.

They did find people in need of help.

Ohio's Auto Safety Stand

The State of Ohio is turning down two of three medical school medical centers for Medicare.
Classical Method

Jacobs Method (1967)

- Area $\times$ Density.


New Formula Counts Crowds

Herbert Jacobs, a lecturer in journalism at the University of California, has devised what he calls "Jacobs' crowd formula." The formula calls for walking off the length and width of a gathering, then adding the two figures. If the crowd is densely packed, this is multiplied by 10; if loosely assembled, by seven.

The average person occupies from six to eight square feet in most crowds, he says. For sit-ins, the figure is 6.7 square feet.

Times Daily, 31 março 1967
Jacobs Method (1967)

- Loose crowds: $1.1 p/m^2$ (10 sqf/p).
- Dense crowds: $2.4 p/m^2$ (4.5 sqf/p).
- Collapsed crowds: $4.3 p/m^2$ (2.5 sqf/p).
Classical Method

Llorenç Badiella - Hotel Majestic - 2010.
Llorenç Badiella - Hotel Majestic - 2010.
Classical Method

Classical Method

Classical Method

Jacobs Method (1967)

- Field observers to determine the occupied area.
- Area split into nearly dense portions of area.
- Perform density measurements at each subarea.

\[
\hat{N} = \sum_i \hat{A}_i \hat{d}_i
\]
Classical Method

Drawbacks (1967)

- Dynamic crowds.
- Early withdrawals or late arrivals.
- Over crowded concentrations where boundaries are exceeded.
- Too high densities.
- Does not provide estimation variability, only point estimates.

Let’s use some statistics!
Measurements obtained by counting all participants in several slices of a given width (1m).

A number of measurements are obtained from several observers allocated systematically along the concentration (or at random).

The length of the concentration has to be estimated, width is irrelevant.
Applicable to:
- Static concentrations.
- Stretched demonstrations
- Moderately dense demonstrations.
- Snapshot at the peak time.
Slice Method

L. Badiella

Attendance estimation for massive demonstrations
Minor oversized concentrations where intersections are occupied are not problematic. All people in the slice are counted.

Sometimes counts are not feasible due to a high density. These values can be imputed and properly weighted.

Preliminary information on the expected attendance can be used to correct the estimate.

Other explanatory variables could be considered.

This method will provide a reliable and valid estimation (with confidence intervals).
Estimation formula (1): Simple scenario

\[ \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \]

\[ \hat{V}(\bar{y}) = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n - 1} \]

Estimation formula (2): With imputed missing data

\[ \bar{y} = \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i} \]

\[ \hat{V}(\bar{y}) = \frac{\sum_{i=1}^{n} w_i (y_i - \bar{y})^2}{df} \]

L. Badiella  
Attendance estimation for massive demonstrations
Estimation formula (3): Stratified sampling with weights

\[\bar{y} = \frac{1}{N} \sum_{h=1}^{H} N_h \bar{y}_h\]

\[\hat{V}(\bar{y}) = \sum_{h=1}^{H} \frac{n_h(1 - f_h)}{n_h - 1} \sum_{i=1}^{n_h} (e_{hi} - e_h)^2\]

\[e_{hi} = \frac{w_{hi}(y_{hi} - \bar{y})}{w_..}\]

Confidence Intervals

\[\bar{y} \pm \sqrt{\hat{V}(\bar{y})} t_{df}^{1-\alpha/2}\]
Attendance estimation for massive demonstrations

<table>
<thead>
<tr>
<th>Font</th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organització</td>
<td>??</td>
<td>2000</td>
<td>??</td>
</tr>
<tr>
<td>Deptm. Interior</td>
<td>1800</td>
<td>1400</td>
<td>540</td>
</tr>
<tr>
<td>Deleg. Gobierno</td>
<td>450-520</td>
<td>550</td>
<td>370</td>
</tr>
<tr>
<td>L. Badiella</td>
<td>800-1000</td>
<td>800-1200</td>
<td>275-420</td>
</tr>
</tbody>
</table>
2014 & 2015 & 2016

Attendance estimation for massive demonstrations

<table>
<thead>
<tr>
<th>Font</th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organització</td>
<td>??</td>
<td>2000</td>
<td>??</td>
</tr>
<tr>
<td>Deptm. Interior</td>
<td>1800</td>
<td>1400</td>
<td>540</td>
</tr>
<tr>
<td>Deleg. Gobierno</td>
<td>450-520</td>
<td>550</td>
<td>370</td>
</tr>
<tr>
<td>L. Badiella</td>
<td>800-1000</td>
<td>800-1200</td>
<td>275-420</td>
</tr>
</tbody>
</table>

Font: Llorenç Badiella - Plaça de les Glòries / Gran Via
### Attendance estimation for massive demonstrations

<table>
<thead>
<tr>
<th></th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total num. of slices</strong></td>
<td>11K</td>
<td>5.2K</td>
<td>1.75K</td>
</tr>
<tr>
<td><strong>Sample size</strong></td>
<td>25</td>
<td>24</td>
<td>15</td>
</tr>
<tr>
<td><strong>Mean Count</strong></td>
<td>82</td>
<td>195</td>
<td>199</td>
</tr>
<tr>
<td><strong>St. Dev.</strong></td>
<td>26</td>
<td>75</td>
<td>74</td>
</tr>
<tr>
<td><strong>Att. Lower Limit</strong></td>
<td>800</td>
<td>850</td>
<td>275</td>
</tr>
<tr>
<td><strong>Att. Upper Limit</strong></td>
<td>1000</td>
<td>1200</td>
<td>420</td>
</tr>
</tbody>
</table>
Conclusions

- The slice method provides reliable and valid estimates.
- Without much effort (25 samples), it provides confidence intervals with a reasonable width.
- “Missing” data and attendance preliminary information can improve estimation.
- The field work is needed to calibrate potential biases of the estimation.
- The field work can be extended to measure complementary characteristics of the attendants, such as gender, age, dress colors, etc.

Q: How do you know there are four elephants in your refrigerator?
A: There’s an empty Mini parked outside.
The slice method provides reliable and valid estimates. Without much effort (25 samples), it provides confidence intervals with a reasonable width.

“Missing” data and attendance preliminar information can improve estimation.

The field work is needed to calibrate potential biases of the estimation.

The field work can be extended to measure complementary characteristics of the attendants, such as gender, age, dress colors, etc.

Q: How do you know there are four elephants in your refrigerator?
A: There's an empty Mini parked outside.
Conclusions

- The slice method provides reliable and valid estimates.
- Without much effort (25 samples), it provides confidence intervals with a reasonable width.
- "Missing" data and attendance preliminary information can improve estimation.
- The field work is needed to calibrate potential biases of the estimation.
- The field work can be extended to measure complementary characteristics of the attendants, such as gender, age, dress colors, etc.
Conclusions

- The slice method provides reliable and valid estimates.
- Without much effort (25 samples), it provides confidence intervals with a reasonable width.
- "Missing" data and attendance preliminary information can improve estimation.
- The field work is needed to calibrate potential biases of the estimation.
Conclusions

- The slice method provides reliable and valid estimates.
- Without much effort (25 samples), it provides confidence intervals with a reasonable width.
- ”Missing” data and attendance preliminary information can improve estimation.
- The field work is needed to calibrate potential biases of the estimation.
- The field work can be extended to measure complementary characteristics of the attendants, such as gender, age, dress colors, etc.
The slice method provides reliable and valid estimates.

Without much effort (25 samples), it provides confidence intervals with a reasonable width.

"Missing" data and attendance preliminary information can improve estimation.

The field work is needed to calibrate potential biases of the estimation.

The field work can be extended to measure complementary characteristics of the attendants, such as gender, age, dress colors, etc.
Conclusions

The slice method provides reliable and valid estimates.

Without much effort (25 samples), it provides confidence intervals with a reasonable width.

"Missing" data and attendance preliminary information can improve estimation.

The field work is needed to calibrate potential biases of the estimation.

The field work can be extended to measure complementary characteristics of the attendants, such as gender, age, dress colors, etc.

Q: How do you know there are four elephants in your refrigerator?
A: There's an empty Mini parked outside.