Proposed statistical analysis to evaluate the qualitative PT of Salmonella serotyping

Marzia Mancin, Lisa Barco, Cristina Saccardin, Antonia Ricci

Marzia Mancin
Statistician
Food Safety Department – Risk Analysis and Public Health Surveillance
Istituto Zooprofilattico Sperimentale delle Venezie
mmancin@izsvenezie.it

Proficiency testing in Salmonella serotyping “AQUA”

• Diagnostic Microbiology
  (Responsable dr. Michela Corro)

• Food Microbiology
  (Responsable dr. Maria Grimaldi)

• Parassitology shellfishes
  (Reference laboratory for Fish health - Responsible dr. Giuseppe Arcangeli)

• Bacteriology and virology of water organisms
  (Reference laboratory for Fish health - Responsible dr. Amedeo Manfrin)

• Virology and Serology
  for avian flu and Newcastle disease
  (Reference laboratory for avian flu and Newcastle disease - Responsible dr. Calogero Terregino)

• Serology and molecular biology
  for bovine and suine diagnostic
  (Responsable dr. Stefano Nardelli)

• Isolation and Salmonella serotyping
  (Reference laboratory for Salmonella – Responsible dr. Antonia Ricci)
• **Isolation:**
  Positive/Negative sample for *Salmonella* spp.

• **Salmonella serotyping:**
  - Identification of somatic antigens
  - Identification of flagellar antigens

Serotype

- S. Typhimurium
- S. Bredney
- S. Virchow
- S. Thompson
- ......

---

• 20 *Salmonella* strains are analysed from each participant

• Routine method is used to analyse the samples

• Results trasmission by e-mail

Standard method is **NOW** available (2014/07/15)
(ISO/TR 6579-3: Guide for serotyping of *Salmonella* spp.)
Proficiency testing in Salmonella serotyping “AQUA”

Eurachem: Berlin, 8 October 2014

Results for each laboratory

<table>
<thead>
<tr>
<th>Strain</th>
<th>Somatic antigen</th>
<th>Flagellar antigens</th>
<th>Serotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3,15,34</td>
<td>γ:1,5</td>
<td>S. Orion</td>
</tr>
<tr>
<td>2</td>
<td>6,7,14</td>
<td>κ:1,5</td>
<td>S. Thompson</td>
</tr>
<tr>
<td>3</td>
<td>6,7,14</td>
<td>r :1,2</td>
<td>S. Virchow</td>
</tr>
<tr>
<td>4</td>
<td>1,4,[5],12</td>
<td>e,h : 1,2</td>
<td>S. Typhimurium</td>
</tr>
<tr>
<td>5</td>
<td>1,9,12</td>
<td>g,m : -</td>
<td>S. Enteritidis</td>
</tr>
<tr>
<td>6</td>
<td>1,9,12</td>
<td>g,m : -</td>
<td>S. Enteritidis</td>
</tr>
<tr>
<td>20</td>
<td>1,4,[5],12</td>
<td>e,h : 1,2</td>
<td>S. Typhimurium</td>
</tr>
</tbody>
</table>

Salmonella serotyping PT evaluation

- **Performance of each participant:**
  Agreement between observed and expected serotypes

- **Performance of overall PT:**
  Agreement among all answers of all participants

Cohen’s K statistics
Cohen's K statistics

K for 2 or more Raters on a 2 or multiple -Level Measurement Scale

- **2 Raters**: results of 2 operators or laboratories, expected vs observed results, results of 2 different methods
- **More Raters**: results of more operators or laboratories, more observed results of one laboratory, results of more different methods

- **2-Level Measurement Scale**: +/-, true/false, presence/absence
- **Multiple-Level Measurement Scale**: an arbitrary number \( q \) (greater than 2) of nominal or ordinal response categories
  - **Ordinal**: absent, mild, moderate, severe; low, medium, high;
  - **Nominal**: different pathologies, diagnosis, categories

K in *Salmonella* serotyping PT

K for 2 Raters on a nominal Multiple-Level Measurement Scale: individual evaluation

- **2 Raters**: observed vs expected results for each laboratory
- **Multiple-Level Measurement Scale**: nominal response categories as serotype for an overall of 20 strains

K for \( p \) Raters on a nominal Multiple-Level Measurement Scale: overall evaluation

- **p Raters**: \( p \) laboratories
- **Multiple-Level Measurement Scale**: nominal response categories as serotype for an overall of 20 strains
Cohen’s K statistics

**Which is the difference between the simple overall percentage of agreement and the K statistics?**

The K statistics **adjusts** the overall percentage of agreement for the chance agreement.

---

**Example 1: % of agreement**

Two doctors analyse 100 x-ray to classify them as pathological or normal:

- **Doctor 1**: he **KNOWS** to distinguish between pathological and normal,
  he analyses the x-ray and decides: 4 pathological, 96 normal
- **Doctor 2**: he **DOESN’T** know to distinguish between pathological and normal,
  he analyses the x-ray and decides: 100 normal

<table>
<thead>
<tr>
<th></th>
<th>Doctor 1</th>
<th>Doctor 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pat. 0</td>
<td>Norm. 0</td>
</tr>
<tr>
<td>Pat.</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Norm.</td>
<td>4</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>96</td>
</tr>
</tbody>
</table>

**Observed agreement:**

\[
\frac{(0+96)}{100} = 0.96 \leftrightarrow 96\%
\]

Is it correct to say that the agreement is equal to 96%?
Example 2: % of agreement

Two doctors analyse 100 x-ray to classify them as pathological or normal:

- **Both the doctors (1 e 2)** flip a coin and decide: 50% pathological and 50% normal in this way:

<table>
<thead>
<tr>
<th></th>
<th>Doctor 1</th>
<th>Norm.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pat.</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>Norm.</td>
<td>25</td>
<td>25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Doctor 2</th>
<th>Norm.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pat.</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>Norm.</td>
<td>25</td>
<td>25</td>
</tr>
</tbody>
</table>

- **Observed agreement:**
  
  \[
  \frac{25+25}{100} = 0.50 \Rightarrow 50%
  \]

**Is it correct to say that the agreement is equal to 50%?**

**No**

A part of agreement depends on chance

Remove the percentage of agreement attributed to chance

---

Cohen’s

**Observed agreement** = OA

**Potential agreement over and above chance** = 1 - EA

\[
K = \frac{(OA - EA)}{(1 - EA)}
\]

K = indicates the proportion of potential agreement, effectively achieved, excluding the chance
Cohen’s K interpretation

<table>
<thead>
<tr>
<th>Maximum disagreement</th>
<th>Observed agreement equal to chance agreement</th>
<th>Maximum agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>0</td>
<td>+1</td>
</tr>
</tbody>
</table>

Qualitative interpretation of K according to Landis and Koch scale

-1: Maximum disagreement
0: Observed agreement equal to chance agreement
+1: Maximum agreement

-1: no agreement
0-0.20: slight agreement
0.21-0.40: fair agreement
0.41-0.60: moderate agreement
0.61-0.80: substantial agreement
0.81-1.0: almost perfect agreement.

Cohen’s K calculation

1. **OA**, proportion of observed concordant results

\[
OA = \frac{25 + 25}{100} = \frac{50}{100} = 0.5
\]

2. **EA**, proportion of expected concordant results

\[
EA = \left( \frac{50 \cdot 50}{100} + \frac{50 \cdot 50}{100} \right) / 100 = 0.5
\]

3. **K**

\[
K = \frac{(OA - EA)}{(1 - EA)} = \frac{0.5 - 0.5}{1 - 0.5} = \frac{0 \neq 0.5}{0.5} = 0
\]

Conclusion: Agreement due to chance
Generalization: Performance of each participant

<table>
<thead>
<tr>
<th>Expected result</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a_{11} )</td>
<td>( a_{12} )</td>
</tr>
<tr>
<td>( a_{21} )</td>
<td>( a_{22} )</td>
</tr>
<tr>
<td>( a_{1} )</td>
<td>( a_{2} )</td>
</tr>
</tbody>
</table>

Generalization: Performance of overall PT

<table>
<thead>
<tr>
<th>Expected result</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a_{11} )</td>
<td>( a_{12} )</td>
</tr>
<tr>
<td>( a_{21} )</td>
<td>( a_{22} )</td>
</tr>
<tr>
<td>( a_{1} )</td>
<td>( a_{2} )</td>
</tr>
</tbody>
</table>

\( K \) for 2 raters on a multiple-Level Measurement Scale

\( K = \frac{\sum_{i=1}^{n} (a_{i} - a_{..})}{a_{..} - \sum_{i=1}^{n} (a_{i} - a_{..})} \)

\( \frac{AO - AA}{1 - AA} \)

\( J_{i} \) for \( r \) raters
K statistic, significance and CI

K statistic

\[ K = \frac{AO - AA}{1 - AA} \]

Significance (p-value)

\[ z = \frac{\hat{K}}{s.e.(\hat{K})} \approx N(0,1) \]

Confidence interval

\[ \hat{K} - z_{1-\alpha/2} \cdot s.e.(\hat{K}) \leq K \leq \hat{K} + z_{1-\alpha/2} \cdot s.e.(\hat{K}) \]

Performance evaluation of *Salmonella* serotyping PT: 2013

<table>
<thead>
<tr>
<th>Typing</th>
<th>K</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>LAB 1</td>
<td>0.89</td>
<td>0.00</td>
</tr>
<tr>
<td>LAB 2</td>
<td>0.83</td>
<td>0.00</td>
</tr>
<tr>
<td>LAB 3</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>LAB 4</td>
<td>0.94</td>
<td>0.00</td>
</tr>
<tr>
<td>LAB 5</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>LAB 6</td>
<td>0.78</td>
<td>0.00</td>
</tr>
<tr>
<td>LAB 7</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>LAB 8</td>
<td>0.89</td>
<td>0.00</td>
</tr>
<tr>
<td>LAB 9</td>
<td>0.94</td>
<td>0.00</td>
</tr>
<tr>
<td>LAB 10</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>LAB 11</td>
<td>0.89</td>
<td>0.00</td>
</tr>
<tr>
<td>LAB 12</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Overall</td>
<td>0.86</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Performance for each laboratory over time

Performance of PT over time
THANK YOU FOR YOUR ATTENTION