Conditioned significance in the test of the difference between proportions for small samples

Francesco Gori
Facoltà di Psicologia, Università di Roma “La Sapienza”

and

Michele Grassi
Dipartimento di Psicologia “G. Kanizsa”, Università di Trieste

Ljubljana, 4 October 2008
The Normal Approximation: Introduction

- Considering two independent samples of sizes \(n_1\) and \(n_2\) extracted under \(H_0\) from a population with two alternatives A and B with probabilities \(\pi\) and \(\kappa\), the amounts of successes \(a_1\) and \(a_2\) can be observed.

- The difference between the proportions, \(p_1 = a_1/n_1\) and \(p_2 = a_2/n_2\), is usually described through the variable:

\[
(p_1 - p_2)/[pq(1/n_1 + 1/n_2)]^{1/2},
\]

whereby

\[
p = (n_1p_1 + n_2p_2)/(n_1 + n_2) = (a_1 + a_2)/(n_1 + n_2),
\]

and \(q = 1 - p\).

**z test:**

The variable (1) uses the maximum likelihood estimator (2) to assess \(\pi\), which is unknown, and makes reference to the standardised Normal distribution.
Dealing with small samples, the procedure uses discrete sample distributions, after establishing the value of $\pi$.

The variable used is $dp = p_1 - p_2$

given $\pi$ and $\kappa = 1 - \pi$ under $H_0$, and the sample extraction with reintroduction, the probability is:

$$ Pr(p_1)Pr(p_2) = \binom{n_1}{a_1} \pi^{a_1} \kappa^{n_1-a_1} \binom{n_2}{a_2} \pi^{a_2} \kappa^{n_2-a_2} \tag{3} $$

the researcher can assess $\pi$ through the average of $p_1$ and $p_2$ produced by the two samples;

in the present case it is not used to standardize the variable $p_1 - p_2$, but it is used only to estimate the parameter $\pi$ of the population under $H_0$. 
Small samples: features of the sample distribution

- Under the null hypothesis ($\pi_1 = \pi_2 = \pi$), the average of the distribution is:
  \[ \mu_1(dp) = E(dp) = E(p_1 - p_2) = E(p_1) - E(p_2) = \pi - \pi = 0 \]

- and the variance is:
  \[ \mu'_2(dp) = E[(dp - 0)^2] = E(dp^2) = \pi\kappa(1/n_1 + 1/n_2) = \sigma_{dp}^2 \]

the asymmetry index:

\[ \gamma_1(dp) = \frac{\mu'_3(dp)}{\sigma_{dp}^{3/2}} = \frac{(\kappa - \pi)(n_2 - n_1)}{[\pi\kappa n_1 n_2 (n_1 + n_2)]^{1/2}} \]  \hfill (4)

the kurtosis index:

\[ \gamma_2(dp) = \frac{\mu'_4(dp)}{\sigma_{dp}^4} - 3 = \frac{(n_1^3 + n_2^3)(1 - 6\pi\kappa)}{n_1 n_2 (n_1 + n_2)^2 \pi \kappa} \]  \hfill (5)
Small samples: features of the sample distribution

The asymmetry index:

\[
\gamma_1(dp) = \frac{\mu_3(dp)}{\sigma_{dp}^{3/2}} = \frac{(\kappa - \pi)(n_2 - n_1)}{[\pi \kappa n_1 n_2(n_1 + n_2)]^{1/2}}
\]

\[6\]

- \(\gamma_1\) is null, as happens for the Normal, for \(\pi = \kappa\)
- the situation improves if \(n_1 = n_2 = n\)
the kurtosis index:

\[ \gamma_2(dp) = \frac{\mu'_4(dp)}{\sigma_{dp}^4} - 3 = \frac{(n_1^3 + n_2^3)(1 - 6\pi\kappa)}{n_1n_2(n_1 + n_2)^2\pi\kappa} \quad (7) \]

- \( \gamma_2 \) is null, as happens for the Normal, for \( \pi \) or \( \kappa = .5 \pm \sqrt{3}/6 \)
- Moreover the same value of \( \pi \) (or \( \kappa \)) vanishing kurtosis would provide the following expression for the asymmetry:
  \[ \frac{(n_2 - n_1)}{[n_1n_2/(n_1 + n_2)^2]^{1/2}} \]
  which is null only if \( n_1 = n_2 \)
- even if samples are of equal size, replacing a discrete distribution with a continuous curve leads to incorrect conclusion.
The distribution is obtained:

considering all possible differences \( dp = p_1 - p_2 \);

ordering with growing \( dp \) and with \( dp \) ranging from -1 to 1;

grouping the probabilities with the same \( dp \)

Sample Extraction:

with reintroduction:

\[
Pr(p_1, p_2) = \frac{n_1}{a_1} \pi^{a_1} \kappa^{n_1-a_1} \frac{n_2}{a_2} \pi^{a_2} \kappa^{n_2-a_2}
\]

without reintroduction:

independent samples

\[
Pr(p_1, p_2) = \frac{N\pi}{a_1} \frac{N\kappa}{n_1-a_1} \frac{N\pi}{a_2} \frac{N\kappa}{n_2-a_2}
\]

non-independent groups

\[
Pr(p_1, p_2 | p_1) = \frac{N\pi}{a_1} \frac{N\kappa}{n_1-a_1} \frac{N\pi - a_1}{a_2} \frac{N\kappa - n_1 + a_1}{n_2 - a_2}
\]
Example 1: Extraction with reintroduction

Two independent groups:

Sample extraction with reintroduction

\[
\begin{align*}
& n_1 = 15; \\& n_2 = 20; \\& a_1 = 8; \\& a_2 = 16 \\
& d_{p_s} = -0.27; \\
& H_0: \pi_1 = \pi_2 = \pi = 0.686 \\
& (\alpha = .05)
\end{align*}
\]

Sample Extraction:

<table>
<thead>
<tr>
<th></th>
<th>(d_p') (f.cum.)</th>
<th>(d_p'') (f.cum.)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>with reintroduction:</td>
<td>-0.283 (0.042)</td>
<td>0.283 (0.039)</td>
<td>0.053</td>
</tr>
<tr>
<td>without reintroduction: independent samples</td>
<td>-0.267 (0.046)</td>
<td>0.267 (0.043)</td>
<td>0.046</td>
</tr>
<tr>
<td>non-independent groups</td>
<td>-0.283 (0.042)</td>
<td>0.283 (0.039)</td>
<td>0.054</td>
</tr>
</tbody>
</table>

Normal: \(z = 1.682; p-value = 0.046\).
Example 2: Extraction without reintroduction

- For some kinds of research, owing to privacy reasons or to limit the advertising of a research initiative, limited masses should be used to extract samples.

- Moreover in some research sector, a treatment carried out on a limited group does not provide the same information given by a sample extracted from a population (or limited mass), which has previously been object of a less accurate and more dispersive treatment.
Example 2: Extraction without reintroduction

An university teacher has a class of N=200 students getting mediocre results and is willing to test a new teaching method.

- the teacher has noticed that if he/she explains the same topics to a limited group of students, in private, everyone seems to understand;

- working with a limited group, the result are better from a quality and quantity points of view compared to results achieved working with the whole class.

- the teacher has to apply the procedure to the whole class and not only to a limited sample of students.
Example 2: Extraction without reintroduction

Two independent samples extracted from a limited mass (N=200):

\[
\begin{align*}
\text{Sample extraction without reintroduction} \\
\{ \\
\quad n_1 &= 15; \quad a_1 = 8 \\
\quad n_2 &= 20; \quad a_2 = 16 \\
\quad d_p &= -0.27; \\
\quad H_0: \pi_1 &= \pi_2 = \pi = 0.686 \\
\quad (\alpha = 0.05)
\end{align*}
\]

Sample Extraction:

<table>
<thead>
<tr>
<th></th>
<th>(d'_p)</th>
<th>(f.cum.)</th>
<th>(d''_p)</th>
<th>(f.cum.)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>with reintroduction:</td>
<td>-0.283</td>
<td>(0.042)</td>
<td>0.283</td>
<td>(0.039)</td>
<td>0.053</td>
</tr>
<tr>
<td>without reintroduction: independent samples</td>
<td>-0.267</td>
<td>(0.046)</td>
<td>0.267</td>
<td>(0.043)</td>
<td>0.046</td>
</tr>
<tr>
<td>non-independent groups</td>
<td>-0.283</td>
<td>(0.042)</td>
<td>0.283</td>
<td>(0.039)</td>
<td>0.054</td>
</tr>
</tbody>
</table>

Normal: \(z = 1.682\); \(p - value = 0.046\).
Example 3: Extraction without reintroduction

Two non-independent samples extracted from a limited mass (N=200):

Sample extraction without reintroduction (first sample not reintroduced)

\[
\begin{align*}
\text{Sample extraction} & \quad \text{without reintroduction} \\
\text{(first sample not reintroduced)} & \quad \begin{cases}
n_1 = 15; & a_1 = 8 \\
n_2 = 20; & a_2 = 16 \\
dp_s = -.27; & \\
H_0: \pi_1 = \pi_2 = \pi = 0.686 & (\alpha = .05)
\end{cases}
\end{align*}
\]

Sample Extraction:

<table>
<thead>
<tr>
<th></th>
<th>dp'</th>
<th>(f.cum.)</th>
<th>dp''</th>
<th>(f.cum.)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>with reintroduction:</td>
<td>-0.283</td>
<td>(0.042)</td>
<td>0.283</td>
<td>(0.039)</td>
<td>0.053</td>
</tr>
<tr>
<td>without reintroduction:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>independent samples</td>
<td>-0.267</td>
<td>(0.046)</td>
<td>0.267</td>
<td>(0.043)</td>
<td>0.046</td>
</tr>
<tr>
<td>non-independent groups</td>
<td>-0.283</td>
<td>(0.042)</td>
<td>0.283</td>
<td>(0.039)</td>
<td>0.054</td>
</tr>
</tbody>
</table>

Normal: \( z = 1.682 \); \( p - valore = 0.046 \).
Simulation with R: results

Example 1

\[
N_1 \leftarrow 1 \times 10^5 \\
b_{11} \leftarrow \text{rbinom} \left(N_1, 15, 0.686\right); n_1 = 15 \\
b_{12} \leftarrow \text{rbinom} \left(N_1, 20, 0.686\right); n_2 = 20 \\
d_{1} = \frac{b_{11}}{n_1} - \frac{b_{12}}{n_2} \\
\text{quantile} \left(d_{1}, c\left(0.042, 1 - 0.039\right)\right) \\
\quad \text{4.2\% \ 96.1\%} \\
\quad -0.2833333 \quad 0.2833333
\]

Example 2

\[
\text{Pop} = \text{c}(\text{rep}(1, 137), \text{rep}(0, 200 - 137)) \\
d_{1} = \text{rep}(0, 1 \times 10^6) \\
\text{for} (i \in 1:1 \times 10^6) \\
\quad s_{1} \leftarrow \text{sample} \left(\text{Pop}, 15, \text{replace} = \text{FALSE}, p = \text{rep}(1/200, 200)\right) \\
\quad s_{2} \leftarrow \text{sample} \left(\text{Pop}, 20, \text{replace} = \text{FALSE}, p = \text{rep}(1/200, 200)\right) \\
\quad p_{1} \leftarrow \text{mean} \left(s_{1}\right); p_{2} \leftarrow \text{mean} \left(s_{2}\right) \\
\quad d_{1}[i] \leftarrow p_{1} - p_{2} \\
\text{quantile} \left(d_{1}, c\left(0.046, 1 - 0.043\right)\right) \\
\quad \text{4.6\% \ 95.7\%} \\
\quad -0.2666667 \quad 0.2666667
\]
The report of p-value

Sample extraction with reintroduction

\[
\begin{align*}
\begin{cases}
 n_1 &= 15; & a_1 &= 10 \\
 n_2 &= 20; & a_2 &= 8 \\
 dp_s &= .27; \\
 H_0: \pi_1 &= \pi_2 = \pi = 0.514 \\
 p-value &= 0.0657; (\alpha = .05)
\end{cases}
\end{align*}
\]

Distribution of p-value as \( \pi \) varies

- **Absolute significance**: the p-value is lower than \( \alpha \) level for every values of \( \pi \)
- **Conditional significance**: results are significant for a subset of values of \( \pi \)
The report of p-value

The case is considered in which a researcher has no information to establish \( \pi \) and chooses it at random:

- in this case the likelihood to reject (or not) the hypothesis \( H_0 \) is provided by the sum of the amplitudes of \( \pi \) intervals.
- for our data, the likelihood to reject the hypothesis \( H_0 \) is provided by the sum of the amplitudes of \( \pi \) intervals, comprised between values close to 0 up to .3 and from .7 up to values close to 1, that is about \(.3 + .3 = .6\).

On the other hand, the researcher is provided with information on the value of \( \pi \):

- having the report or the graph, can observe whether the value that he/she chose is supported by a more or less large subset comprising it, that is to say whether his/her evaluation of \( \pi \) is part of, or even better is the centre of a subset, in which the test is significant or not.

The researcher can reflect upon the proposed theory and model, above all if the test is significant in a subset of values of \( \pi \) deemed senseless.
contacts and utilities

- Francesco Gori: francesco.gori@uniroma1.it
- Michele Grassi: grassim@psico.units.it

Software DPT.exe and DPT package for R environment:
- p-value
  - with reintroduction
  - without reintroduction (surveyed fraction or population given by the user)
- tables of critical values
  - for different sample sizes
  - and subsets of values of $\pi$
- confidence interval of $\pi$
- report of p-value in function of $\pi$
  - continuous case: DPT package provides the extreme points of the subset of $\pi$ through an immediate solution.