Funnel plot for institutional comparison: the funnelcompar command

Silvia Forni    Rosa Gini
silvia.forni@arsanita.toscana.it
rosa.gini@arsanita.toscana.it

Agenzia regionale di sanità
della Toscana

September 10, 2009
1. Funnel plot for institutional comparison

2. Some statistics
   - Underlying test
   - Exact vs approximated control limits

3. The funnelcompar command

4. Some examples
1. Funnel plot for institutional comparison

2. Some statistics
   - Underlying test
   - Exact vs approximated control limits

3. The `funnelcompar` command

4. Some examples
Background

Funnel plots for comparing institutional performance

David J. Spiegelhalter*

SUMMARY

‘Funnel plots’ are recommended as a graphical aid for institutional comparisons, in which an estimate of an underlying quantity is plotted against an interpretable measure of its precision. ‘Control limits’ form a funnel around the target outcome, in a close analogy to standard Shewhart control charts. Examples are given for comparing proportions and changes in rates, assessing association between outcome and volume of cases, and dealing with over-dispersion due to unmeasured risk factors. We conclude that funnel plots are flexible, attractively simple, and avoid spurious ranking of institutions into ‘league tables’. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS: control charts; outliers; over-dispersion; institutional profiling; ranking
Background

- **Quantitative indicators** are increasingly used to monitor health care providers.
- Interpretation of those indicators is often open to anyone (patients, journalists, politicians, civil servants and managers).
- It is crucial that indicators are both accurate and presented in a way that does not result in unfair criticism or unjustified praise.
Classical presentation: *league tables*

- Imply the existence of **ranking** between institutions
- Implicitly support the idea that some of them are worse/better than other
Statistical Process Control methods: key principles

- Variation, to be expected in any process or system, can be divided into:
  - **Common cause variation**: expected in a stable process
  - **Special cause variation**: unexpected, due to systematic deviation

- Limits between these two categories can be set using SPC methods

- Funnel plots:
  - All institutions are part of a single system and perform at the same level
  - Observed differences can never be completely eliminated and are explained by chance (*common cause variation*).
  - If observed variation exceed that expected, *special-cause variation* exists and requires further explanation to identify its cause.
Funnel Plot

- **Scatterplot** of observed indicators against a measure of its precision, typically the sample size
- **Horizontal line** at a target level, typically the group average
- **Control Limits** at 95% (≈ 2SD) and 99.8% (≈ 3SD) levels, that narrow as the sample size gets bigger

Association of Public Health Observatories in UK developed analytical tools in Excel for producing funnel plot
1. Funnel plot for institutional comparison

2. Some statistics
   - Underlying test
   - Exact vs approximated control limits

3. The `funnelcompar` command

4. Some examples
A funnel plot has four components:

- An *indicator* $Y$.
- A *target* $\theta$ which specifies the desired expectation for institutions considered “in control”.
- A *precision* parameter $N$ determining the accuracy with which the indicator is being measured. Select a $N$ directly interpretable, e.g., the denominator for rates and means.
- *Control limits* for a $p$-value, computed assuming $Y$ has a known distribution (normal, binomial, Poisson) with parameters $(\theta, \sigma)$. 
From a purely statistical point of view, funnel plot is a graphical representation testing whether each value $Y_i$ belongs to the known distribution with given parameters.

The formal test of significance:

$$H_0 : Y_i = \theta$$

$$H_1 : Y_i \neq \theta$$

$$Z = \frac{Y_i - \theta}{(\sigma/\sqrt{N})}$$

- test failed 99.8%: alarm
- test satisfied: in control
- test failed 95%: alert
Control limits

In cases of discrete distributions there are two possibilities for drawing control limits as functions of $N$

- a normal approximation:
  \[ y_p(N) = \theta \pm z_p \frac{\sigma}{\sqrt{N}} \]

- an “exact” formula
  \[ y_p(N) = \frac{r(p,N,\theta) - \alpha}{N} \]

where $r(p,N,\theta)$ and $\alpha$ are defined in the following slides.
In the case of binomial distribution:

- \( r(p,N,\theta) \) is the inverse to the cumulative binomial distribution with parameters \((\theta, N)\) at level \(p\). The definition Spiegelhalter refers to is as follows:\(^1\) if \( F(\theta, N) \) is the cumulative distribution function, ie \( F(\theta, N)(k) \) is the probability of observing \( k \) or fewer successes in \( N \) trials when the probability of a success on one trial is \( \theta \),\(^2\) then \( r_p = r(p,N,\theta) \) is the smallest integer such that

\[
P(R \leq r_p) = F(\theta, N)(r_p) > p
\]

- \( \alpha \) is a continuity adjustment coefficient

\[
\alpha = \frac{F(\theta, N)(r_p) - p}{F(\theta, N)(r_p - 1) - p}
\]

---

\(^1\)Beware that the Stata function \texttt{invbinomial()} is \textit{not} defined this way.

\(^2\)The Stata function \texttt{binomial(N,k,\theta)} computes \( F(\theta, N)(k) \).
In the case of Poisson distribution:

- $r(p,N,\theta)$ is the inverse to the cumulative Poisson distribution with parameter $M = \theta N$ at level $p$. The definition Spiegelhalter refers to is as follows: if $F_M$ is the cumulative distribution function, i.e., $F_M(k)$ is the probability of observing $k$ or or fewer outcomes that are distributed Poisson with mean $M$, then $r_p = r(p,N,\theta)$ is the smallest integer such that

$$P(R \leq r_p) = F_M(r_p) > p$$

- $\alpha$ is a continuity adjustment coefficient

$$\alpha = \frac{F_M(r_p) - p}{F_M(r_p - 1) - p}$$

---

3 Beware that the Stata function invpoisson() is not defined this way.

4 The Stata function poisson(M,k) computes $F_M(k)$. 
Example 1: binomial, $\theta=1\%$

- Does it make sense to test a $1\%$ of cases with $N < 100$?
- For $N \geq 100$ the two pairs of curves almost coincide

---

From invbinomom2(), probability .01
Example 2: binomial, $\theta=20\%$

- For $N < 100$ very similar curves, approximated upper bounds conservative
- For $N > 100$ the two pairs of curves almost coincide
Example 3: binomial, $\theta=50\%$

- For $N < 100$ very similar curves, approximated upper bounds conservative
- For $N > 100$ the two pairs of curves almost coincide

From `invbinomom2()`, probability .5
Example 4: Poisson, $\theta=1\%$

- Does it make sense to test a 1% of cases with $N<100$?
- For $N>100$ the two pairs of curves almost coincide.
Example 5: Poisson, $\theta=50\%$

The two pairs of curves almost coincide.

From `invpoisson2()`, rate .5
Example 6: Poisson, $\theta=1$ (SMR)

The two pairs of curves visibly coincide.
Conclusion for using exact vs approximated test

- Whenever the sample size is more than 100, the approximated test is almost superimposed to the exact test
- Consider if it makes sense to use exact test
1. Funnel plot for institutional comparison

2. Some statistics
   - Underlying test
   - Exact vs approximated control limits

3. The funnelcompar command

4. Some examples
funnelcompar value pop unit [sdvalue],
[continuous/binomial/poisson]
[ext_stand() ext_sd() noweight smr ]
[constant()]
[contours() exact]

marking options
other options
funnelcompar value pop unit [sdvalue]

- *value* contains the values of the indicator.
- *pop* contains the sample size (precision parameter)
- *unit* contains an identifier of the units
- *sdvalue* contains the standard deviations of indicators (optionally, if the continuous option is also specified)
Distribution

Users must specify a distribution among:

- \textit{normal}: option \texttt{cont}
- \textit{binomial}: option \texttt{binom}
- \textit{Poisson}: option \texttt{poiss}
Parameters: $\theta$

$\theta$ can be obtained as:

- weighted mean of value with weights $pop$ (default)
- non weighted mean of value if the $noweight$ option is specified
- external value specified by users with the option $ext\_stand()$
Parameters: \( \sigma \)

- Binomial distribution: \( \sigma = \sqrt{\theta(1 - \theta)} \)
- Poisson distribution: \( \sigma = \sqrt{\theta} \)
- Normal distribution:
  - weighted mean of sdvalue with weights pop (default)
  - non-weighted mean of sdvalue if the noweight option is specified
  - external value specified by users with the option ext_sd()
The smr option

- `smr` option can be specified only with `poisson` option:
- `value` are assumed to be indirectly standardised rates
- `pop` contains the expected number of events
- $\theta$ is assumed to be 1
Constant

- The `constant()` option specifies whether the values of the indicators are multiplied by a constant term, for instance `constant(100)` must be specified if the values are percentages.
Curves

- `contours()`: specifies significance levels at which control limits are set (as a percentage).
- Default `contours()` are set at 5% and .2% levels, that is a confidence of 95% and 99.8% respectively.
- For example if `contours(5)` is specified only the curve corresponding to a test with 5% of significance is drawn.
- For discrete distributions if the exact option is specified, the exact contours are drawn. As a default the normal approximation is used.
1 Funnel plot for institutional comparison

2 Some statistics
   - Underlying test
   - Exact vs approximated control limits

3 The funnelcompar command

4 Some examples
Percentages, internal target, units out-of-control marked

funnelcompar
measure pop unit, binom const(100)
markup marklow
Funnel plot for institutional comparison
Some statistics
The funnelcompar command
Some examples

Percentages, no-weighted internal target

funnelcompar
measure pop unit,
binom const(100)
noweight
funnelcompar measure pop unit, poisson
const(1000) ext_stand(15)
markcond(type = 1)
legendmarkcond(Type A)
colormarkcond(blue)
optionsmarkcond(msymbol(S))
twowayopts(yline(23, lcolor(green)))
Means, internal target, unit type marked

funnelcompar measure pop unit sd, cont const(1) markcond(type=1) legendmarkcond(Type A) colormarkcond(blue) optionsmarkcond(msymbol(S)) markcond1(type = 2) ...markcond2(type=3) ...
Standardized Incidence Rates, one unit marked

funnelcompar smr exp unit, poisson smr markunit(5 "your unit") legendopts(placement(se) row(1))
References

Spiegelhalter DJ

Spiegelhalter DJ

Spiegelhalter DJ

Julian Flowers
Acknowledgements

- We thank Neil Shephard and Paul Silcocks for valuable discussion.
- Our routine is heavily based on confunnel by Tom Palmer.
- Many programming tricks were stolen from eclplot and other routines by Roger Newson.
Thanks for your attention!