

From: FlowJo Support - Maciej Simm <flowjo@treestar.com>
Subject: Fwd: Question about correlations
Date: August 28, 2008 10:15:17 AM PDT
To: Maciej Simm <simmm@treestar.com>
 7 Attachments, 68.0 KB

how to do phi corrs in FlowJo:

Starting with this table, aaron is right, we need four inputs:

		Attribute 1	
Attribute 2	Yes	No	
Yes	<i>a</i>	<i>b</i>	
No	<i>c</i>	<i>d</i>	

let's convert abcd's to quadrants, while keeping in mind that above table shows attribute 1 in reverse of how we'd plot it in a Graph Window, so:

a=UR (double pos)
 b=UL
 c=LR
 d=LL (double neg)

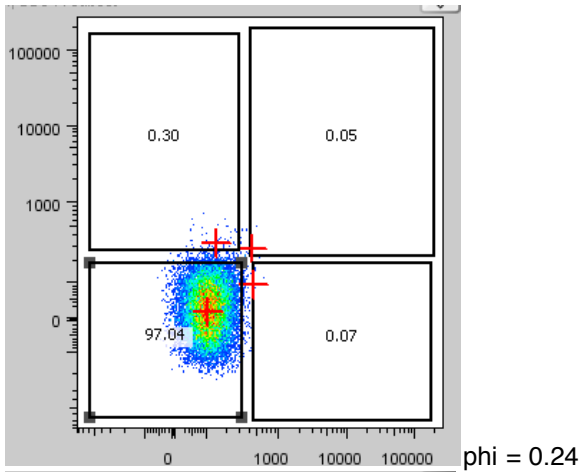
Make these 4 gates (or one set of quads) and drag them from the Workspace window to the Table window. Name them UR/UL etc, or a b c d, according to the scheme above. Either way, we'll combine these 4 gates into this expression using the "formula" tool in the table editor:

$$\phi = \frac{ad - bc}{\sqrt{(a + b)(c + d)(a + c)(b + d)}}$$

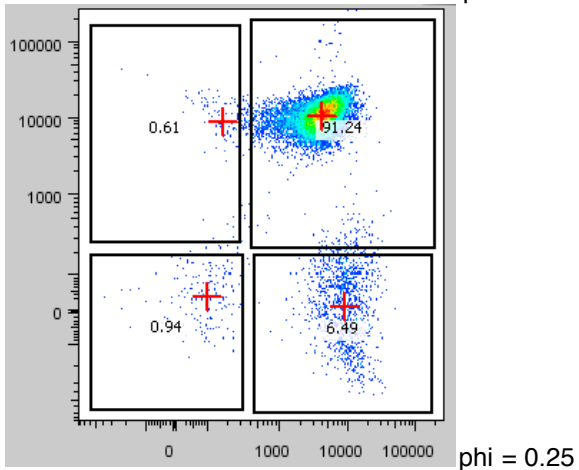
Which looks a bit like this:

Col.	Name	Description
1	UL	Statistic: Freq. of Parent (gate: 'FSC-A, SSC-A subset/UL')
2	UR	Statistic: Freq. of Parent (gate: 'FSC-A, SSC-A subset/UR')
3	LL	Statistic: Freq. of Parent (gate: 'FSC-A, SSC-A subset/LL')
4	LR	Statistic: Freq. of Parent (gate: 'FSC-A, SSC-A subset/LR')
5	numerator	Formula: ((<Column "UR">*<Column "LL">)-(<Column "UL">*<Column "LR">))*100
6	denominator	Formula: 100*(sqrt((<Column "UR">+<Column "UL">)*(<Column "LL">+<Column "LR">))*(<Column "UR">+<Column "LR">)*(<Column "UL">+<Column "LL">)))
7	phi	Formula: ((<Column "UR">*<Column "LL">)-(<Column "UL">*<Column "LR">))/sqrt((<Column "UR">+<Column "UL">)*(<Column "LL">+<Column "LR">)*(<Column "UR">+<Column "LR">)*(<Column "UL">+<Column "LL">)))

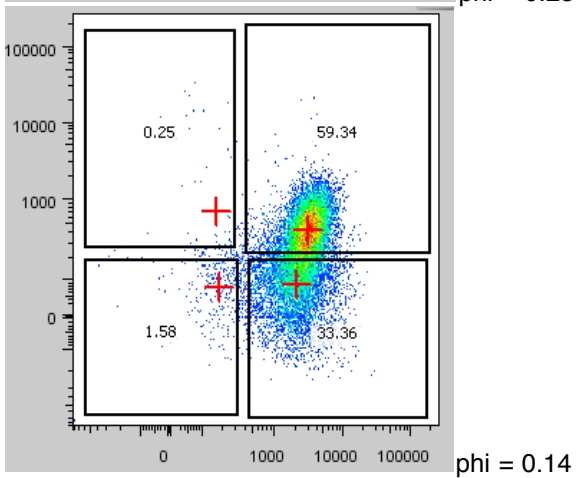
and test it on some data:



phi = 0.24



phi = 0.25



phi = 0.14

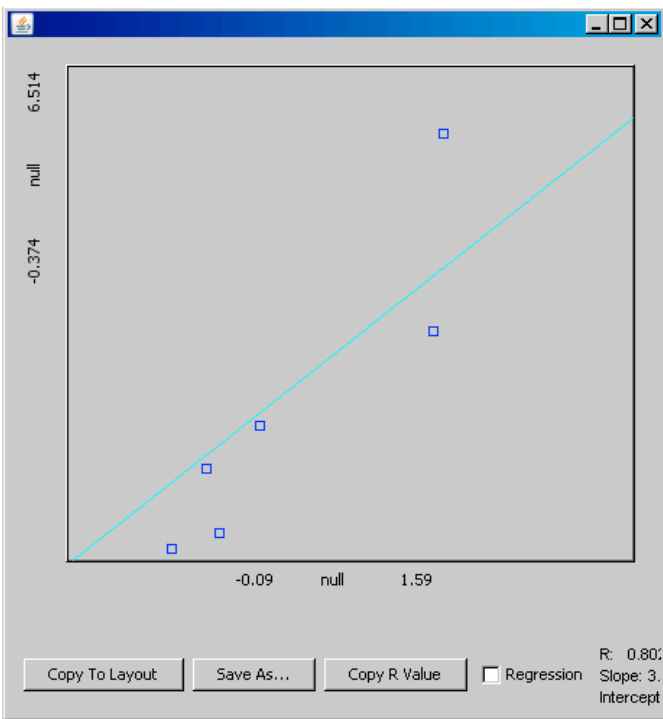
this matches expected results, which are that the more events exist in (UL & LR) the lower the phi. The more evenly split the events are among LL & UR, the higher the phi, so a hypothetical distribution like..

0 50
50 0

makes a phi of 1.

oh, and table editor's correlation can be used if you keep numerator and denominator as two different formulae:

Col.	Name	Description
1	UL	Statistic: Freq. of Parent (gate: 'FSC-A, SSC-A subset/UL')
2	UR	Statistic: Freq. of Parent (gate: 'FSC-A, SSC-A subset/UR')
3	LL	Statistic: Freq. of Parent (gate: 'FSC-A, SSC-A subset/LL')
4	LR	Statistic: Freq. of Parent (gate: 'FSC-A, SSC-A subset/LR')
5	numerator	Formula: ((<Column "UR">*<Column "LL">)-(<Column "UL">*<Column "LR">))*100
6	denominator	Formula: 100*(sqrt((<Column "UR">+<Column "UL">)*(<Column "LL">+<Column "LR">)*(<Column "UR">+<Column "LR">)*(<Column "UL">+<Column "LL">))))
7	phi	Formula: ((<Column "UR">*<Column "LL">)-(<Column "UL">*<Column "LR">))/(sqrt((<Column "UR">+<Column "UL">)*(<Column "LL">+<Column "LR">)*(<Column "UR">+<Column "LR">)*(<Column "UL">+<Column "LL">))))



the last correlation is probably not pertinent, since it's a inter-sample correlation, whereas ofir is looking to establish intra-sample correlation between two parameters.

ms