THE GINI COEFFICIENT

The Gini coefficient is a statistic which measures the ability of a scorecard or a characteristic to rank order risk. A Gini value of 0% means that the characteristic cannot distinguish good from bad cases, e..g.

	Goods	Bads	Bad Rate
Phone	80%	80%	15%
No Phone	20%	20%	15%

A Gini value of 100% means that a characteristic/scorecard distinguishes perfectly.

	Goods	Bads	Bad Rate
Phone	100%	0%	0%
No Phone	0%	100%	100%

A typical credit scorecard has a Gini coefficient of 40-60%. Behaviour scorecards have values of 70-80%. A very powerful characteristic can have a Gini coefficient of 25%.

To calculate Gini values, assume that one has good and bad accounts rank ordered by score with the score sufficiently finely graded such as that there is only one case per score. The essential notion is that of a "flip". A flip is a transposition of consecutive good and bad accounts.

The Gini coefficient is the percentage of flips required to reach the rank ordering from a random assignment of goods and bads by score (i.e. with Gini = 0). The example below illustrates this, with the accounts to be flipped underlined.

EXAMPLE

(b = bad, g = good, cases ranked in increasing score order)

(1) b b g b g b b g g g b g g g g (the starting order, ranked by score)

2) b b b g g b b g g g b g g g g
3) b b b g b g b g g g b g g g g
4) b b b b g g b g g g b g g g g
5) b b b b g b g g g g b g g g g
6) b b b b b g g g g g b g g g g
7) b b b b b g g g g b g g g g g
8) b b b b b g g g b g g g g g g
9) b b b b b g g b g g g g g g g
10) b b b b b g b g g g g g g g g
b b b b b b g g g g g g g g g

- Therefore there are 10 flips to reach the ideal state from the actual (where the ideal state is that shown after flip 10).

- A scorecard that had no effect would rank the goods and bads randomly, therefore on average each bad would have to flip with half of the goods. Therefore to get from the random state to the ideal state would take 6 x 9/2 flips (there are 6 bad and 9 goods) = 27 flips.

- Therefore, to reach state (1) from the random state would take 27-10 flips = 17 flips out of a maximum of 27. The Gini coefficient is the percentage of such flips = 17/27 = 63%.

Where there is more than one account per score interval, or where one is dealing with a discrete characteristic a refinement to the calculation is necessary, but the principle is the same.

Assume we have a characteristic with n attributes (or alternatively a scorecard with n scores or n score intervals). Let $A_1,\ldots, A_n$ be the attributes. Let G_i be the probability of being attribute i given that the account is good, and B_i be the probability of being attribute i given that the account is bad. Let G be the vector of G_i s and B be the verctor of B_i s, e.g.:

	G	**B**
Tenant`	**.2**	**.5**
LWP	**.1**	**.1**
Owner	**.7**	**.4**

Let U be the upper triangular matrix. Then the Gini coefficient γ is given by the formula.

$$\gamma = G\,(U^T - U)\,B$$

E.G., from the values above, $U = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$ &

$$\gamma = [.2,.1,.7] \begin{bmatrix} 0 & -1 & -1 \\ 1 & 0 & -1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} .5 \\ .1 \\ .4 \end{bmatrix}$$

$$= [.8,.5,-3] \begin{bmatrix} .5 \\ .1 \\ .4 \end{bmatrix} = 0.33 \quad \text{(or 33\%)}$$

There are two alternative calculation methods. The first uses the logic in the following SAS program :

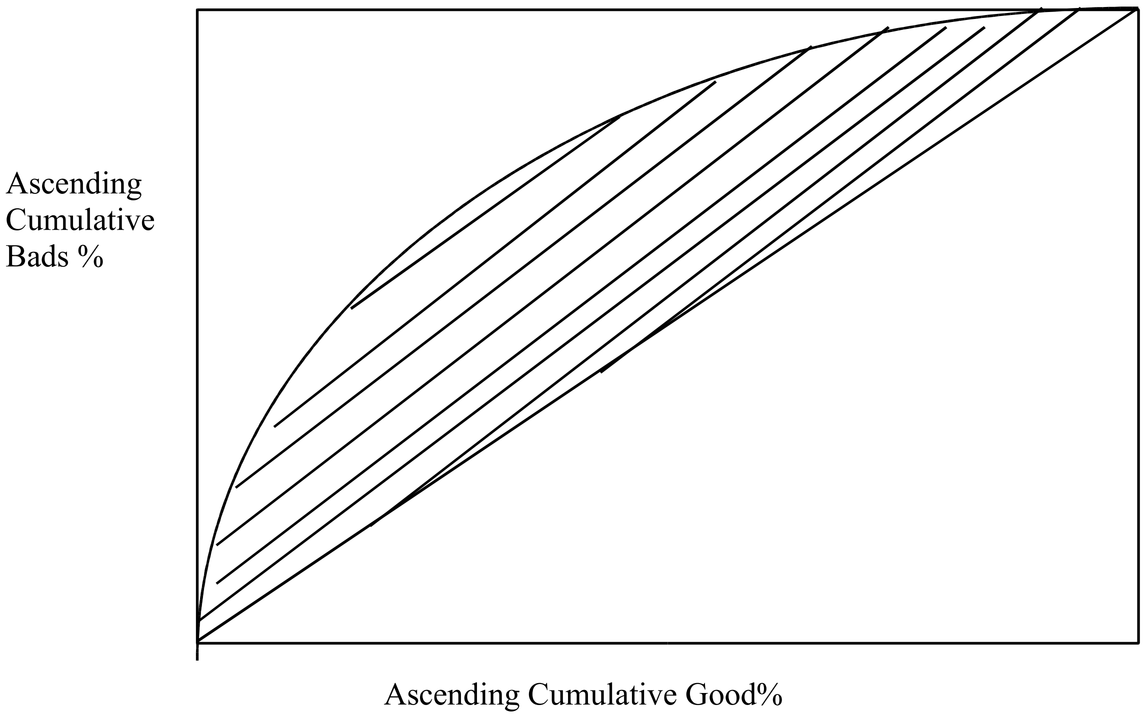
```
INPUT GOODS BADS;
BETWEEN + 2 X CUMGOODS X BADS;
WITHIN + GOODS * BADS;
CUMGOODS + GOODS;
CUMBADS + BADS;
GINI = 100 x [ 1 -  BETWEEN+WITHIN      ]
                   ---------------------
                   CUMGOODS x CUMBADS
```

which is a derivative of the formula below:

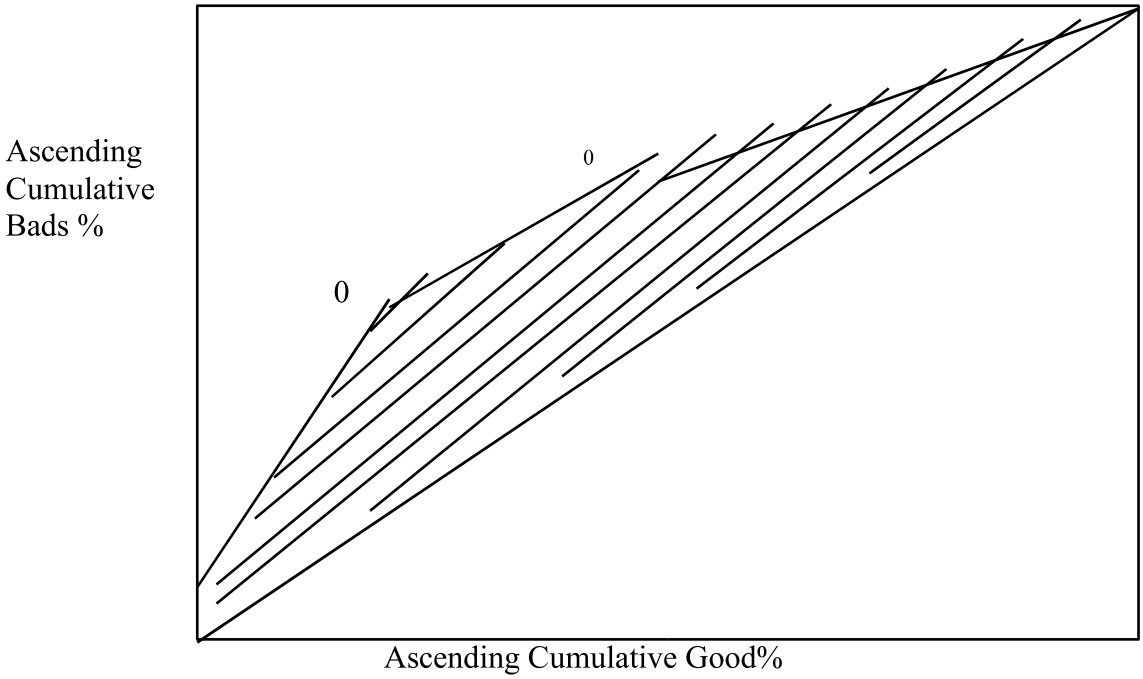
$$\frac{\gamma}{100} = 1 - \left[\frac{\sum_{r=2}^{n} \left(B_r \sum_{s=1}^{r-1} G_s\right) + \frac{1}{2} \sum_{r=1}^{n} G_r B_r}{\frac{1}{2} \sum_{r=1}^{n} G_r \sum_{r=1}^{n} B_r} \right]$$

The other is a direct calculation of the shaded area in the following graph expressed as a percentage of the triangle above the diagonal.

Scorecard



Characteristic with 3 attributes (represented by the 3 lines)



Key points regarding the Gini are :

- it is unaffected by uniformly weighting either the goods or the bads.
- it is very simple to calculate it for a binary characteristic e.g.

	G	B
Tenant	.3	.5
Owner	**.7**	**.5**

The Gini is just .7 - .5 = 0.2 (i.e. 20%)

- When calculating it for a continuous characteristic, one usually assumes that the rank ordering is the natural one. When calculating it for a discrete characteristic, one assumes that the rank ordering is a logical one if one is concerned with the fine-classed values, and a reducing bad rate order if one is concerned with course-classed values.

- Negative Gini values are possible if the characteristic rank orders in the opposite way to expectation.

- The Gini coefficient has certain logical qualities. For instance, if an attribute is split into two such that each component has the same bad rate then the Gini value will not change. This desirable feature is not mirrored by other power measures such as “Divergence”.

Example Pascal Gini and R-Squared Program

```
Program GINI;
Uses
   CRT;
Var
   infile    : string;
   f,g       : text;
   n         : integer;
   cumgoods, cumbads, goods, bads        : real;
   within, between, mss, tss, giny, rsq : real;

Begin
assign(g,'statlist.lst'); rewrite(g);
Writeln(g,'Program GINI'); Writeln(g);
infile :='dat1.txt';
Write (g,'File : '); Writeln(g,infile); writeln(g);
Assign(f,infile); Reset(f);
cumgoods := 0; cumbads := 0; within := 0; between := 0;
n := 0; mss := 0; tss := 0;
While not eof(f) do
   Begin
   Readln(f, goods, bads);
   n := n + 1;
   mss := mss + (goods * bads) / (goods + bads);
   between := between + 2 * bads * cumgoods;
   within := within + goods * bads;
   cumgoods := cumgoods + goods;
   cumbads := cumbads + bads;
   End;
tss := (cumgoods * cumbads) / (cumgoods + cumbads);
rsq := 100 * (1 - mss / tss);
giny := 100 * (1 - (within + between) / (cumgoods * cumbads));
Write (g,'Cumulative Goods : ');
Writeln (g,cumgoods:8:2);
Write (g,'Cumulative Bads  : ');
Writeln(g,cumbads:8:2);
Writeln(g);
Write(g,'Gini %       : ');
Writeln(g,giny:8:2);
Write(g,'R-Squared % : ');
Writeln(g,rsq:8:2);
close(f);
close(g);
End.
```

Example Quick Basic Gini and R-Squared Program

```
'Program to calculate GINI coefficients
COLOR  0,  7,  8
CLS
PRINT "Program GINI" : PRINT
INPUT "Input File : " ; infile$
OPEN "I", 1, infile$
cumgoods = 0
cumbads = 0
within = 0
between = 0
n = 0
DO    WHILE NOT EOF (1)
      INPUT #1, goods, bads
      n = n + 1
```

```
      between = between + 2 * bads * cumgoods
      within = within + goods * bads
      cumgoods = cumgoods + goods
      cumbads = cumbads + bads
LOOP
PRINT “Cumulative Goods  :  “ ;  cumgoods
PRINT “Cumulative Bads     : “ ;  cumbads
PRINT
gini  =  100 * (1  -  (within  +  between)  /  (cumgoods  *  cumbads))
PRINT “Gini  :  “;  : PRINT USING “####.##“; gini
END
```

Alan Lucas
Rhino Risk Ltd.
Alanlucas@rhinorisk.com
www.rhinorisk.com