Information Value Calculation for Binary Dependent Variable

- Page 2 shows Information Value Calculated by info_bin.sas
  - The ranges are determined by even accounts (about 10% each). The WOE* has zigzag trend (so does ODDS).
  - The information value is higher than that on page 3/4 because the noise has not been smoothed out.
- Page 3 shows Information Value Calculated by info_binn.sas with &PEAKS=1
  - &PEAKS=1 (default) allows WOE to have at most ONE peak (maximum/minimum value) along positive side (or negative side). Range >2.5 – 3.0 is a peak (minimum).
  - The significant digit was set to 2 and the ranges look better than those by info_bin.sas.
- Page 4 shows Information Value Calculated by info_binn.sas with &PEAKS=0
  - &PEAKS=0 allows WOE to have no peak at all along positive side (or negative side), which means that WOE must be monotonic along either positive side or negative side.
  - The significant digit was set to 3 and the ranges look better than those by info_bin.sas.

Note: Information Value of page 2 is greater than Information Value of page 3 since the noise is not smoothed out, and
Information Value of page 3 is greater than Information Value of page 4 since &PEAKS=0 will force WOE to be monotonic.

*: WOE = Weight Of Evidence, see definition on page 2.

Information Value Calculated by info_bin.sas

Weight Of Evidence (WOE) is the Natural Log of (First% / Second%): If the percentage of First is greater than the percentage of Second, then WOE is negative. Otherwise is positive or zero (when equal)

Log(9.3%/14.1%) = -0.42

Ln(9.3%/14.1%) = -0.42 (Excel)

Variable Name

Variable's Label (if any)

Characteristic: ATTR0007
Label: ()

Attribute	First Principal #	First Principal %	Second Principal #	Second Principal %	WT of Evid	Each Cell Odds
Missing	15	0.3	2	0.3	-0.10	7.500
<0	20	0.4	11	1.8	-1.52	1.818
Zero	116	2.3	9	1.5	0.44	12.89
>0-1.05	325	6.5	53	8.8	-0.31	6.132
>1.05-1.82	466	9.3	85	14.1	-0.42	5.482
>1.82-2.37	458	9.1	72	11.9	-0.27	6.361
>2.37-2.90	454	9.0	89	14.8	-0.49	5.101
>2.90-3.43	495	9.9	79	13.1	-0.28	6.266
>3.43-3.91	449	8.9	72	11.9	-0.29	6.236
>3.91-4.35	518	10.3	42	7.0	0.39	12.33
>4.35-4.75	630	12.6	44	7.3	0.54	14.32
>4.75-5.05	521	10.4	18	3.0	1.25	28.94
>5.05-High	551	11.0	27	4.5	0.90	20.41
All	5018	100.0	603	100.0		

Over all-population Odds: 8.322
Info Value of ATTR0007: 29.81

It's the population with the dependent variable &BAD = 0 (sample case BAD1=0)

It is equal to (# of First) / (# of second)

15/2=7.5

454 / 5018 = 9% (0.09)

42 / 603 = 7% (0.07)

Sums of the Columns

5018 / 603 = 8.322

It's the population with the dependent variable &BAD = 1 (sample case BAD1=1)

Information Value (IV) is the Sum of each range's sub-IV which is defined as

(First% - Second%) *WOE

=

(First% - Second%)* LOG(First% / Second%)

Information Value Calculated by info_binn.sas with &PEAKS=1

Weight Of Evidence (WOE) is the Natural Log of (First% / Second%): If the percentage of First is greater than the percentage of Second, then WOE is negative. Otherwise is positive or zero (when equal)

Log(9.1%/13.9%) = -0.43

Ln(9.1%/13.9%) = -0.43 (Excel)

Variable Name

Variable's Label (if any)

Characteristic: (ATTR0007)
Label:()

Attribute	First Principal #	First Principal %	Second Principal #	Second Principal %	WT of Evid	Each Cell Odds
Missing	15	0.3	2	0.3	-0.10	7.500
<0	20	0.4	11	1.8	-1.52	1.818
Zero	116	2.3	9	1.5	0.44	12.89
>0.0000-2.5000	1361	27.1	228	37.8	-0.33	5.969
>2.5000-3.0000	456	9.1	84	13.9	-0.43	5.429
>3.0000-3.5000	411	8.2	74	12.3	-0.40	5.554
>3.5000-3.9000	419	8.3	64	10.6	-0.24	6.547
>3.9000-4.3000	479	9.5	39	6.5	0.39	12.28
>4.3000-4.7000	566	11.3	41	6.8	0.51	13.80
>4.7000-High	1175	23.4	51	8.5	1.02	23.04
All	5018	100.0	603	100.0		

Over all-population Odds: 8.322
Info Value of ATTR0007: 29.04

It's the population with the dependent variable &BAD = 0 (sample case BAD1=0)

It is equal to (# of First) / (# of second)

15/2=7.5

&SIGDIGIT=2

419 / 5018 = 8.3% (0.083)

51 / 603 = 8.5% (0.085)

Sums of the Columns

5018 / 603 = 8.322

A peak is a local minimum or maximum.

-0.43 is less than both -0.33 and -0.4

Also can see from the last column

Each Cell Odds

Information Value (IV) is the Sum of each range's sub-IV which is defined as

(First% - Second%) *WOE

=

(First% - Second%)* LOG(First% / Second%)

It's the population with the dependent variable &BAD = 1 (sample case BAD1=1)

Information Value Calculated by info_binn.sas with &PEAKS=0

Weight Of Evidence (WOE) is the Natural Log of (First% / Second%): If the percentage of First is greater than the percentage of Second, then WOE is negative. Otherwise is positive or zero (when equal)

Log(9.3%/12.4%) = -0.29

Ln(9.3%/12.4%) = -0.29 (Excel)

Variable Name

Variable's Label (if any)

It's the population with the dependent variable &BAD = 0 (sample case BAD1=0)

It is equal to (# of First) / (# of second)

15/2=7.5

&SIGDIGIT=3

452 / 5018 = 9% (0.09)

57 / 603 = 9.5% (0.095)

Sums of the Columns

5018 / 603 = 8.322

A peak is a local minimum or maximum.

&PEAKS=0 means WOE must be monotonic on positive side and monotonic on negative side.

Zero and Missing won't count

Information Value (IV) is the Sum of each range's sub-IV which is defined as

(First% - Second%) *WOE

=

(First% - Second%)* LOG(First% / Second%)

It's the population with the dependent variable &BAD = 1 (sample case BAD1=1)

Characteristic: (ATTR0007)
Label:()

Attribute	First Principal #	First Principal %	Second Principal #	Second Principal %	WT of Evid	Each Cell Odds
Missing	15	0.3	2	0.3	-0.10	7.500
<0	20	0.4	11	1.8	-1.52	1.818
Zero	116	2.3	9	1.5	0.44	12.89
>0.0000-2.9100	1730	34.5	303	50.2	-0.38	5.710
>2.9100-3.4300	468	9.3	75	12.4	-0.29	6.240
>3.4300-3.8300	417	8.3	65	10.8	-0.26	6.415
>3.8300-4.2200	452	9.0	42	7.0	0.26	10.76
>4.2200-4.6200	539	10.7	39	6.5	0.51	13.82
>4.6200-High	1261	25.1	57	9.5	0.98	22.12
All	5018	100.0	603	100.0		

Over all-population Odds: 8.322
Info Value of ATTR0007: 28.03