

Information Value Calculation for Continuous Dependent Variable

- Page 2 shows Information Value Calculated by info_con.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_conn.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 >130 – 150 is a peak (maximum).
 - 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_conn.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 2 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_con.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)

$\text{Log}(9.4\%/3.8\%) = 0.91$
 $\text{Ln}(9.4\%/3.8\%) = 0.91 \text{ (Excel)}$

Variable's Label (if any) Variable Name

It's the number of accounts with the dependent variable &YDEPORIG ^= . (sample case REVENUE ^=.)

It is equal to (# of First) / (# of second)
711/872=0.815

Characteristic: (ATTR0005)
Label: (C)

Attribute	First Principal #	First Principal %	Second Principal #	Second Principal %	WT of Evid	Each Cell Odds
>0-70	711	11.9	872	10.2	0.15	0.815
>70-90	546	9.1	809	9.4	-0.04	0.675
>90-120	638	10.6	701	8.2	0.26	0.910
>120-140	643	10.7	371	4.3	0.91	1.731
>140-160	562	9.4	323	3.8	0.91	1.740
>160-185	627	10.5	653	7.6	0.32	0.960
>185-210	549	9.2	484	5.7	0.48	1.133
>210-290	520	8.7	849	9.9	-0.13	0.613
>290-620	605	10.1	1512	17.6	-0.56	0.400
>620-High	592	9.9	1998	23.3	-0.86	0.296
All	5993	100.0	8573	100.0		
over all-population odds:					0.699	
Info value of ATTR0005:					30.33	

549 / 5993 = 9.2% (0.092)

1998 / 8573 = 23.3%

Sums of the Columns

5993 / 8573 = 0.699

Sum of the dependent variable &YDEPORIG (sample case: revenue) divided by the overall mean of &YDEPORIG which was calculated after it has been trimmed at 2%(default)

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

$$\begin{aligned}
 &(\text{First}\% - \text{Second}\%) * \text{WOE} \\
 &= \\
 &(\text{First}\% - \text{Second}\%) * \text{LOG}(\text{First}\% / \text{Second}\%)
 \end{aligned}$$

Information Value Calculated by info_conn.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)

$$\text{Log}(10.8\%/5.2\%) = 0.74$$

$$\text{Ln}(10.8\%/5.2\%) = 0.74 \text{ (Excel)}$$

Variable's Label (if any)

Variable Name

Characteristic: (ATTR0005)
Label: ()

Attribute	First Principal #	%	Second Principal #	%	WT of Evid	Each Cell Odds
>0-85	1113	18.6	939	15.7	0.17	1.186
>85-110	550	9.2	413	6.9	0.29	1.332
>110-130	517	8.6	328	5.5	0.45	1.575
>130-150	649	10.8	311	5.2	0.74	2.089
>150-220	1621	27.0	1032	17.2	0.45	1.570
>220-370	455	7.6	747	12.5	-0.50	0.609
>370-670	553	9.2	1080	18.0	-0.67	0.512
>670-High	535	8.9	1144	19.1	-0.76	0.468
All	5993	100.0	5993	100.0		

over all population odds: 1.000
Info value of ATTR0005: 27.19

It's the number of accounts with the dependent variable &YDEPORIG ^= . (sample case REVENUE ^=.)

&SIGDIGIT=2

$$455 / 5993 = 7.6\% (0.076)$$

$$1144 / 5993 = 19.1\%$$

Sums of the Columns

$$5993 / 5993 = 1 \text{ (always)}$$

It is equal to (# of First) / (# of second)
1113/939=1.186

A peak is a local minimum or maximum.
0.74 is greater than both 0.45 and 0.45
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

$$(\text{First\%} - \text{Second\%}) * \text{WOE}$$

$$=$$

$$(\text{First\%} - \text{Second\%}) * \text{LOG}(\text{First\%} / \text{Second\%})$$

First of all, &YDEPORIG is trimmed at 2%(default).
Sum of the dependent variable &YDEPORIG (sample case: revenue) divided by the overall mean of &YDEPORIG.

Note: This treatment forces the totals to be equal no matter what the scale of the dependent variable is.

Information Value Calculated by info_conn.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 positive. Otherwise is negative or zero (when equal)

$\text{Log}(7.6\%/12.5\%) = -0.5$
 $\text{Ln}(7.6\%/12.5\%) = -0.5$ (Excel)

Variable Name
 Variable's Label (if any)

Characteristic: (ATTR0005)
 Label: ()

Attribute	First Principal #	First Principal %	Second Principal #	Second Principal %	WT of Evid	Each Cell Odds
>0-220	4450	74.3	3023	50.4	0.39	1.472
>220-370	455	7.6	747	12.5	-0.50	0.609
>370-670	553	9.2	1080	18.0	-0.67	0.512
>670-High	535	8.9	1144	19.1	-0.76	0.468
All	5993	100.0	5993	100.0		

over all-population Odds: 1.000
 Info Value of ATTR0005: 25.22

It's the number of accounts with the dependent variable &YDEPORIG ^= . (sample case REVENUE ^=.)

&SIGDIGIT=2

455 / 5993 = 7.6% (0.076)

1144 / 5993 = 19.1%

Sums of the Columns

5993 / 5993 = 1 (always)

It is equal to (# of First) / (# of second)
 4450/3023=1.472

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

$$\text{Sub-IV} = (\text{First\%} - \text{Second\%}) * \text{WOE}$$

$$\text{Info Value} = \sum (\text{Sub-IV}) = \sum (\text{First\%} - \text{Second\%}) * \text{LOG}(\text{First\%} / \text{Second\%})$$

First of all, &YDEPORIG is trimmed at 2%(default). Sum of the dependent variable &YDEPORIG (sample case: revenue) divided by the overall mean of &YDEPORIG.

Note: This treatment forces the totals to be equal no matter what the scale of the dependent variable is.