Predictive Modeling Using Logistic Regression

Step-by-Step Instructions

This document is accompanied by the following Excel Template

> IntegrityM Predictive Modeling Using Logistic Regression in Excel Template.xlsx

Two datasets are used to run predictive modeling based on prior information:

- Training dataset This dataset includes both historical and current data with distinction of the outcomes coded 1 for "Yes" and 0 for "No". Steps 1 to 7 use the dataset to assess the relative importance of indicators/variables to the outcome (Weights) and to determine the strength of their association with the outcome (P-values).
- Scoring dataset new data on individuals/entities/IDs used to compute the probability of outcome.

<u>Step 1:</u>

Input your analytical file; in our example, we downloaded data from the Office of Inspector General List of Excluded Individuals and Entities (LEIE) and from the Centers for Medicare and Medicaid Services (CMS) Public Use File (PUF). Columns A to G of the spreadsheet "Predictive Model" display the data.

The analytical file includes columns A to G. Column A displays the provider identifier and columns B to G show summary information on the variables per provider inputted to the analysis, including the response variable informing if the provider was excluded or not (coded "1' or "0" respectively). The working dataset consists of information stored in cells B2 to G92.

Steps 2 and 3:

Copy values from column H (RANDBETWEEN Values) to column I (P (E)) – this is necessary because of the volatile nature of the formula in column H, which will values changed every time the spreadsheet is refreshed. The rest of the columns in Step 3 will be computed automatically by the formulas embedded in them (Column J (Odds Ratio), Colum K (If Odds Ratio < 0.1 set it to 0.1) and Column L (Log (Odds Ratio))

- The Excel function RANDBETWEEN assigns random probability numbers to the providers according to their original classification ("1" or "0") drawing from the range 0.50 to 0.99 for providers coded 1 (excluded) and from the range 0.01 to 0.49 for providers coded 0 (active in the health care program)
- Note that the RANDBETWEEN function requires as input values integers (numbers) that define the bottom and the top values from which a random number will be drawn. The values will be stored in column H, "RANDBETWEEN Values". The assigned values need to be transformed to probability numbers by dividing the function output number by 100 according to the provider exclusion status and the following Excel operation:
- for excluded providers, =RANDBERWEEN(50, 99)/100;
- ➢ for non-excluded (active) providers, =RANDBERWEEN(1, 49)/100;



<u>Step 4</u>:

- ➢ Step 4.1:
 - Run the Linear Regression Model by using the Data Analysis tool of Excel as shown in the screenshot below to obtain the Initial weights (coefficients) of the variables/indicators (in our example, 5 variables). The regression input Y Range (response variable) is the "Log(Odds)", column L; and the five indicators in Input X Range are all values in "PMT per Bene", "Services per Bene", "Average Age of Beneficiaries", "Beneficiary cc diab percent" and "Average HCC Risk Score of Beneficiary" columns C to G. The Regression Model results will generate a new tab labeled in our example "Step 4 Reg Initial Values".
- Step 4.2:
 - Copy the coefficients (weights) in column B from the regression model output to the Coefficients Table (in our example, the table includes cells T3 to T8 in column T of the spreadsheet "Predictive Model").



INTEGRITY MANAGEMENT SERVICES

Regression	Carlo Anno Lana	? ×
Input Input <u>Y</u> Range: Input <u>X</u> Range: ✓ Labels ✓ Con <u>f</u> idence Level: 95	\$L\$2:\$L\$92 \$C\$2:\$G\$92 Constant is <u>Z</u> ero %	OK Cancel <u>H</u> elp
Output options Output Range: New Worksheet <u>P</u> ly: New <u>W</u> orkbook Residuals		
<u>Residuals</u> Standardized Residuals Normal Probability <u>N</u> ormal Probability Plots	Resi <u>d</u> ual Plots	

<u>Step 5:</u>

Update the formula in Column N (variable labeled L) if you have more or less than the 5 variables that we have in our example. This formula includes the initial values of the weights from the Coefficients Table (Column T in our example) multiplied by the corresponding indicators (columns C to G in our example). Changes in the set of summation terms in the cases of 4 or 6 variables would be

- 4 variables, formula =T\$3 + T\$4*C4 + T\$5*D4 + T\$6*E4 + T\$7*F4 + T\$8*G4
- 6 variables, formula = T\$3 + T\$4*C4 + T\$5*D4 + T\$6*E4 + T\$7*F4 + T\$8*G4+T\$9*H4



Note: the mathematical constant "e" (=2.7183 in cell T10 in our example) is required to compute the results in column O (labeled eL = "e to power L").

The values of column P (P(X) = eL/(1 + eL)) and column Q ($Y^*In[P(X)] + (1 - Y)^*In[1 - P(X)]$) will be updated automatically (where Y = 1 for excluded providers and Y = 0 for active providers) by the embedded formulas.

<u>Step 6</u>:

- Step 6.1:
 - Update the summation of the Log (Maximum Likelihood) if you have more rows than our example:
 - If you have 100 rows in your data rather than the 90 rows of our example, change the formula from =SUM(Q3:Q92) to =SUM(Q3:Q102)
- > Step 6.2:
 - Run the Excel Solver tool to obtain the set of final weights that define the relative importance of the indicators/variables to the outcome. The Excel Solver will work on the set of initial weights (previously generated by the logistic regression) to update the Coefficients Table with a set of final weights that maximizes the likelihood of obtaining the data (outcome variable and indicators) actually observed. The Excel Solver will generate an output tab that is labeled "Step 6 - Solver Final Values" in our example.

From Access			perties	Clear Reapply				+ Data Analysis
Fro	om Other Existing ources * Connections	Refresh	Z Sort Filter	Advanced Columns Fi	Solver Parameters			X
	ternal Data	Connectic						
					Set Objective:	\$U\$2		
U2 *	$\therefore \qquad \checkmark \qquad f_x$							······································
A	В	R	S	Т	To: 💿 Max	🔘 Mi <u>n</u> 🔍 Valu	e Of: 0	
1			Step 5		By Changing Variable	Colley		
					\$T\$3,\$T\$4,\$T\$5,\$T\$6			
2 ID	EXCLUDED			4 4077	3133,3134,3133,3130	0,9197,9190		
3 ID_710840 4 ID_320133	1		ntercept MT per Bene	4.4977	Subject to the Constr-	aints:		
5 ID_728367	1		ervices per Bene	-0.1550			*	Add
6 ID_136308	1		verage Age of Beneficiario					
7 ID_180319	1		eneficiary cc diab percent					⊆hange
8 ID_607721	1		verage HCC Risk Score of					Delete
9 ID_338287	1							
10 ID_932418	1	e		2.7183				Reset All
11 ID_774841	1	Lo	og(Maximum Likelihood)	-36.9574				
12 ID_524888	1						Ÿ	Load/Save
13 ID_238313	1				Ma <u>k</u> e Unconstrai	ned Variables Non-Negative		
14 ID_331177	1				Select a Solving Meth	od: GRG Nonlinear	•	Options
15 ID_270247	1					onomina		Oguons
16 ID_662302	1				Solving Method			
17 ID_279521 18 ID_186590	1					linear engine for Solver Problems t oblems, and select the Evolutionary		
19 ID_930119	1				Ior infeat Solver Pro	solutional activity the Evolutional y	engine for power problems th	
20 ID_376426	1							
21 ID_257242	0				Help]	Solve	Close
22 ID_543001	0				Дер	J	Zoive	Cigse
	lytical File 5 Variab	es Step 3 -	Reg Initial Values Ste	o 6 - Solver Final Values	Step 7 - Cov (+)	4		
POINT CALCULATE	, , , , , , , , , , , , , , , , , , , ,				, , , ,		III 🛙	III
		- 20						4 20 014
🌝 💇	🧿 😂 🔤		S 🖌 🕅			R S 🚳 👎 📴	💿 🗖 🦁 💿 🎽 🔒	12/1/2016



rom Access	2 Connections			Clear					E	. 6		_			← []	
	Properties	Z↓ <u>A Z</u> Z↓ Sort	I Filter	Reapply	Text to	Flash	Remove	Data	Consoli	ata W	1	Polatio			Ungroup :	Eubtotal
rom Text Sources - Connections All -	Edit Links	Â↓ Son	Filter	🍹 Advanced	Columns			s Validatior			ilysis *	Relatio	nsnips	- Toroup	•	Subtotal
Get External Data Co	onnections	5	Sort & Filte	Br				Data	Tools						Outline	
\cdot : \times \checkmark f_x																
A B C	D	E	F	G	Н		I	J	К	L		М		N	0	Р
Objective Cell (Max)			_													
	Original Value															
\$T\$:Log(Maximum Likelihood)	-36.9574	-17.5649	_													
Variable Cells																
	Original Value	Final Value	Integer	-				Г								
\$T\$:Intercept	4.4977	17.7796	-	_												
\$T\$4PMT per Bene	0.0057		Contin	-												
\$T\$! Services per Bene	-0.1550	-0.9372		-												
\$T\$! Services per Bene \$T\$! Average Age of Beneficiaries	-0.1550 -0.0694		Contin	_												
\$T\${ Average Age of Beneficiaries \$T\$; beneficiary cc diab percent	-0.0694 1.7622	-0.9372 -0.2324 11.1129	Contin Contin Contin	-												
STSt Average Age of Beneficiaries	-0.0694	-0.9372 -0.2324	Contin Contin Contin													
\$T\$(Average Age of Beneficiaries \$T\$; beneficiary cc diab percent	-0.0694 1.7622	-0.9372 -0.2324 11.1129	Contin Contin Contin													
\$T\$(Average Age of Beneficiaries \$T\$: beneficiary cc diab percent \$T\$(Average HCC Risk Score of Benefi	-0.0694 1.7622	-0.9372 -0.2324 11.1129	Contin Contin Contin													
\$T\$I Average Age of Beneficiaries \$T\$I:beneficiary cc diab percent \$T\$I Average HCC Risk Score of Benefi Constraints	-0.0694 1.7622 -1.0249	-0.9372 -0.2324 11.1129	Contin Contin Contin													
\$T\$(Average Age of Beneficiaries \$T\$:beneficiary cc diab percent \$T\$\Average HCC Risk Score of Benefi Constraints	-0.0694 1.7622	-0.9372 -0.2324 11.1129	Contin Contin Contin	-												
\$T\$(Average Age of Beneficiaries \$T\$:beneficiary cc diab percent \$T\$\Average HCC Risk Score of Benefi Constraints	-0.0694 1.7622 -1.0249	-0.9372 -0.2324 11.1129	Contin Contin Contin													
\$T\$(Average Age of Beneficiaries \$T\$:beneficiary cc diab percent \$T\$\Average HCC Risk Score of Benefi Constraints	-0.0694 1.7622 -1.0249	-0.9372 -0.2324 11.1129	Contin Contin Contin	-												
\$T\$I Average Age of Beneficiaries \$T\$I:beneficiary cc diab percent \$T\$I Average HCC Risk Score of Benefi Constraints	-0.0694 1.7622 -1.0249	-0.9372 -0.2324 11.1129	Contin Contin Contin	-												
\$T\$I Average Age of Beneficiaries \$T\$I:beneficiary cc diab percent \$T\$I Average HCC Risk Score of Benefi Constraints	-0.0694 1.7622 -1.0249	-0.9372 -0.2324 11.1129	Contin Contin Contin	-												
\$T\$(Average Age of Beneficiaries \$T\$: beneficiary cc diab percent \$T\$! Average HCC Risk Score of Benefi Constraints	-0.0694 1.7622 -1.0249 NONE	-0.9372 -0.2324 11.1129 -8.8086	Contin Contin Contin Contin	-												
\$T\$(Average Age of Beneficiaries \$T\$:beneficiary cc diab percent \$T\$(Average HCC Risk Score of Benefi Constraints	-0.0694 1.7622 -1.0249	-0.9372 -0.2324 11.1129 -8.8086	Contin Contin Contin Contin	- - - - - - - - - - - - - - - - - - -	al Value:	s St	ер 7 - Со	\ (+)								

Note: the signs of the regression coefficients indicate whether additional units of the associated variables increase (positive sign) or decrease (negative sign) the probability that the analyzed provider will reach the target outcome (joining the OIG Exclusion List in this example).

For example, the average payment per beneficiary (positively associated with the probability of exclusion) for excluded providers \$342.52 as opposed to \$214.26 for non-excluded (active) providers; on the other hand, the average risk score per beneficiary (negatively associated with the probability of exclusion) is 1.1632 for excluded providers in contrast with 1.7361 for non-excluded providers.

<u>Step 7:</u>

Steps 7.1 to 7.7 determine the strength of the association (p-values) of each Indicator/variable (explanatory variable) with the outcome variable (the probability of being excluded from the Medicare program in our example) by estimating the covariance matrix of the model.

- ➢ <u>Step 7.1:</u>
 - Construct an Excel table having a row per provider and a column per variable *without including the provider identifier*; add to the table a column of "1s" (intercept). In our example the table includes columns A to F of the spreadsheet "Step 7 Covariance Matrix" (intercept plus five variables "PMT per Bene", "Services per Bene", "Average Age of Beneficiaries", "Beneficiary cc diab percent" and "Average HCC Risk Score of Beneficiary").
- ➢ <u>Step 7.2</u>:



Copy values from column P (labeled P(X) = eL/ (1 + eL)) from tab "Predictive Model" to column G (labeled P(X) = eL/ (1 + eL)) in tab "Step 7 - Covariance Matrix". Values in column H (labeled (1 - P(X))) and column I (labeled P(X)*(1 - P(X))) will be updated automatically by the embedded formulas.

Note: From steps 7.3 to 7.6, remember to highlight the destination cells (range of cells) BEFORE completing the formula with Shift-Control-Enter

Note: The destination of the transpose operation needs to be highlighted (cells B97 to CM102 in our example) before the re-scaled table (cells K4 to P93) is transposed. Operations with matrices (tables) in Excel are explained in "Notes on Matrix Operations in Excel" by Ronald Larsen, http://www.eng.auburn.edu/~clemept/CEANALYSIS SPRING2011/matrixoperations_notes.pdf

- ➢ <u>Step 7.3</u>:
 - Multiply every row of the table created in step 7.1 (cells A4 to G93 in our example) by the values in column I (P(X)*(1 P(X))), to generate a re-scaled table stored in cells K4 to P93. This multiplication is accomplished in row 4 by the Excel function "=I4*(A4:F4)", which is copied and pasted to all rows from 5 to 93.

Note: Remember to highlight the destination cells (range of cells) BEFORE completing the formula with Shift-Control-Enter

- Step 7.4:
 - Transpose the re-scaled table stored in cells K4 to P93 using the Excel matrix function "TRANSPOSE(re-scaled table)" to generate the transposed table stored in cells B97 to CM102 in our example.

Note: Remember to highlight the destination cells (range of cells) BEFORE completing the formula with Shift-Control-Enter

Cut Calibri 11	× A A	= = 🐴 🦻	R W Wrap Text	Number		E X AutoSu	[™] [•] ? ★	-
Paste ▼	<u>ð</u> - <u>A</u> -		Merge & Center	- \$ - % • .00 .00	Conditional Format as Cell Formatting * Table * Styles *	Insert Delete Format	Sort & Find & Filter - Select -	
Clipboard 5 Font	5	Alic	anment	G Number G		Cells	Editing	
	{=TRANSPOS		,				carony	
A	в	c	D	E	F	G	н	
15			ep 7.4					
6 TRANSPOSE W = {P(X)*(1 P(X))*X}'	1	2	3	4	5	6	7	
7 Intercept	0.1499	0.1662	0.0000	0.0003	0.1912	0.0207	0.1392	
8 PMT per Bene	27.8888	51.0835	0.0048	0.1412	33.5494	11.9098	16.8284	
9 Services per Bene	0.5663	0.9196	0.0001	0.0021	0.6979	0.2773	0.3052	
00 Average Age of Beneficiaries	10.7961	12.1340	0.0004	0.0185	11.8572	1.5075	9.6020	
01 beneficiary cc diab percent	0.0375	0.1247	0.0000	0.0001	0.0708	0.0072	0.0654	
02 Average HCC Risk Score of Benefi	0.1439	0.2581	0.0000	0.0003	0.2479	0.0211	0.1716	
03								1
04			ep 7.5					
D5 PRODUCT (W*X)						rage HCC Risk Score of Benefi		
06 Intercept	5.5065	1225.3803	25.5291	378.4397	2.0458	6.4677		
07 PMT per Bene	1225.3803		6905.6104	85825.0651	469.0428	1507.0698		
08 Services per Bene 09 Average Age of Beneficiaries	25.5291 378.4397	6905.6104 85825.0651	151.7947 1787.7186	1787.7186 26189.4344	9.9388 141.7404	30.6930 444.1432		
10 beneficiary cc diab percent	2.0458	469.0428	9,9388	141.7404	0.8834	2.5105		
11 Average HCC Risk Score of Benefi	6.4677	1507.0698	30,6930	444.1432	2.5105	7.9981		
12	0.4077	1307.0058	30.0330	444.1452	2.5105	7.9901		
13		St	ep 7.6					
14 COVARIANCE MATRIX								
15 INVERSE OF (X'*W)	Intercept	PMT per Bene	Services per Benear	ge Age of Beneficiaries	beneficiary cc diab percent	rage HCC Risk Score of Benefi		
16 Intercept	59.3103	0.0596	-1.4418	-0.7794	16.3653	-15.5222		
17 PMT per Bene	0.0596	0.0002	-0.0049	-0.0008	0.0308	-0.0270		
Analytical File 5 Variables	Step 3 - R	leg Initial Values	Step 6 - Solver	Final Values Step 7	- Cov (+) 🕴 4			Þ
EADY					AVERAGE: 3 0433	COUNT: 540 SUM: 1643.3692	□ □	- + 100



- ➢ <u>Step 7.5</u>:
 - Multiply the transpose of the re-scaled table, computed in the previous step and stored in cells B97 to CM102, by the original (not re-scaled) table stored in cells A4 to F93 using the Excel matrix function "MMULT(transpose of the re-scaled table, original table)" to create a product table stored in cells B106 to G111.

Note: Remember to highlight the destination cells (range of cells) BEFORE completing the formula with Shift-Control-Enter.

aste → Format Painter → Format Painter → Format Painter	A A	= = <mark>▲</mark> ≫ •	R W Wrap Text	General ~ r ∗ \$ ∗ % ۶ (€.0,.00 .00 →.0	Conditional Format as Cell Formatting * Table * Styles *	Insert Delete Format	ZI
Clipboard 5 Font		Al	ignment	R Number R	Styles	Cells	Editing
3106 • : $\times \checkmark f_x$	{=MMULT(B9	97:CM102,A4:F	93)}				
A	В	С	D	E	F	G	н
04		S	tep 7.5				
05 PRODUCT (W*X)	Intercept	PMT per Bene	Services per Benea	ge Age of Beneficiaries	beneficiary cc diab percent?	rage HCC Risk Score of Benefi	
06 Intercept	5.5065	1225.3803	25.5291	378.4397	2.0458	6.4677	
07 PMT per Bene	1225.3803		6905.6104	85825.0651	469.0428	1507.0698	
08 Services per Bene	25.5291	6905.6104	151.7947	1787.7186	9.9388	30.6930	
09 Average Age of Beneficiaries	378.4397	85825.0651	1787.7186	26189.4344	141.7404	444.1432	
L0 beneficiary cc diab percent	2.0458	469.0428	9.9388	141.7404	0.8834	2.5105	
11 Average HCC Risk Score of Benefi	6.4677	1507.0698	30.6930	444.1432	2.5105	7.9981	
12							
13		5	itep 7.6				
		DAAT	C		have distance with the second	rage HCC Risk Score of Benefi	
15 INVERSE OF (X'*W) 16 Intercept	59.3103	0.0596	-1.4418	e Age of Beneficiaries -0.7794	16.3653	-15.5222	
17 PMT per Bene	0.0596	0.0596	-1.4418	-0.7794	0.0308	-15.5222 -0.0270	
18 Services per Bene	-1.4418	-0.0049	0.1892	0.0008	-0.9667	0.7623	
19 Average Age of Beneficiaries	-0.7794	-0.0049	0.0162	0.0102	-0.2241	0.1688	
20 beneficiary cc hypert percent	16.3653	0.0308	-0.9667	-0.2241	17.6422	-8.4198	
21 Average HCC Risk Score of Benefi	-15.5222	-0.0270	0.7623	0.1688	-8.4198	8.1025	
22							
23		S	itep 7.7				
24			-				
25	Coef	VAR	STD DEV	Wald	p-value		
26 Intercept	17.7796	59.3103	7.7013	5.3299	0.0210		
Analytical File 5 Variables	Step 3 - R	eg Initial Values	Step 6 - Solver	Final Values Step 7	- Cov (+) : (+)		

<u>Step 7.6</u>:

Compute the inverse of the table created in the step 7.5 using the Excel matrix function "MINVERSE(product table)" to create the covariance matrix, which is stored in cells B116 to G121. The highlighted main diagonal values as shown below are the variances of the variables to be used in Step 7.7.

Note: Remember to highlight the destination cells (range of cells) BEFORE completing the formula with Shift-Control-Enter



Integrity Management Services

Past	Copy -	1 · A A · <u>8</u> · <u>A</u> ·		Wrap Text	Number ▼ er \$ ▼ % \$ \$ €.0 .00 .00 .00 .00	Conditional Format as Cell Formatting * Table * Styles	Insert Delete Format	Cost Q. Find Q.	
	Clipboard 🖡 Font	5	Ali	gnment	G Number G	Styles	Cells	Editing	^
B11	$16 \cdot : \times \checkmark f_x$	{=MINVERSE	(B106:G111)}						*
	А	В	С	D	E	F	G	н	^
113			S	tep 7.6					
114	COVARIANCE MATRIX								
115	INVERSE OF (X'*W)	Intercept	PMT per Bene	Services per Bene	age Age of Beneficiaries	beneficiary cc diab percent	rage HCC Risk Score of Benef	i	
116	Intercept	59.3103	0.0596	-1.4418	-0.7794	16.3653	-15.5222	2	
117	PMT per Bene	0.0596	0.0002	-0.0049	-0.0008	0.0308	-0.0270)	
118	Services per Bene	-1.4418	-0.0049	0.1892	0.0162	-0.9667	0.7623	8	
119	Average Age of Beneficiaries	-0.7794	-0.0008	0.0162	0.0110	-0.2241	0.1688	3	
120	beneficiary cc hypert percent	16.3653	0.0308	-0.9667	-0.2241	17.6422	-8.4198	3	
121	Average HCC Risk Score of Benefi	-15.5222	-0.0270	0.7623	0.1688	-8.4198	8.1025	5	
122								Ţ	



Step 7.7:

- Copy the coefficients (weights) computed by the Excel Solver from the cells T3 to T8 (in our example) from tab "Predictive Model" to the cells B126 to B131 of tab "Step 7 Covariance Matrix". The final weights are also found in the Excel Solver output tab.
- Copy the variances of the variables from the main diagonal of the covariance matrix cells B116, C117, D118, E119, F120 and G121 - to cells C126 to C131 in column C of tab "Step 7 -Covariance Matrix".

Note: the next steps will calculate new values automatically using the weights and variances inputted in the previous steps.

- The standard deviations are computed using the Excel function "SQRT(variance)" applied to values in cells C126 to C131 and store the results in cells D126 to D131 in our example.
- The Wald Chi-square statistic for each variable is computed as the squared value of the ratio (coefficient/standard deviation) and stored it in cells E126 to 131 under the label "Wald" in our example.
- The strength of the association "p-value" is computed using the Excel function "CHISQ.DIST.RT(Wald,1)", where the second input to the function - the number 1 - is the number of "degrees of freedom" in this case. The p-values are stored in cells F126 to F131.

Note: The P-value of each variable assesses the strength of the association between the variable and the outcome. The lower the p-value the stronger the association between variables. The standard rule is to consider p-values <= 5% (0.05) as indicative of statistically significant association. In our example, all variables were found to be strongly associated with the outcome.

	Cut Calibri • 1	1 • A A =	= 🚽 🇞 •	Wrap Text	General *	- 🖳 💓 🚽
Past	te ≪ Format Painter ■ I U - 🗄					Conditional Format as Cell Formatting * Table * Styles
	Clipboard 🔹 Font	Fa	Alig	nment	rs Number rs	Styles
F13	3 • i × √ ƒx					
1	A	В	С	D	E	F
23			Ste	ep 7.7		
24						
25		Coef	VAR	STD DEV	Wald	p-value
26	Intercept	17.7796	59.3103	7.7013	=(B126/D126)^2	<pre>=CHISQ.DIST.RT(E126,1)</pre>
27	PMT per Bene	0.0359	0.0002	0.0126	8.1835	0.0043
28	Services per Bene	-0.9372	0.1892	0.4350	4.6418	0.031
29	Average Age of Beneficiaries	-0.2324	0.0110	0.1050	4.9017	0.026
30	beneficiary cc diab percent	11.1129	17.6422	4.2003	7.0001	0.0083
31	Average HCC Risk Score of Benefi	-8.8086	8.1025	2.8465	9.5763	0.0020

The screenshot below reproduces the results of the predictive model analysis using the same data but implemented with the statistical package SAS, which generates exactly the same results as the ones obtained using Excel.

Analysis of Maximum Likelihood Estimates								
Parameter	DF	Estimate	Standard	Wald	Pr > ChiSq			
			Error (Chi-Square				
Intercept	1	17.7796	7.7013	5.3299	0.0210			
PMT_per_Bene	1	0.0359	0.0126	8.1835	0.0042			
Services_per_Bene	1	-0.9372	0.4350	4.6418	0.0312			
Average_Age_of_Benef	1	-0.2324	0.1050	4.9017	0.0268			
beneficiary_cc_diab_	1	11.1129	4.2003	7.0001	0.0082			
Average HCC Risk Sco	1	-8.8086	2.8465	9.5763	0.0020			

<u>Step 8:</u>



Steps 8 to 10 use a new dataset of active (unclassified as excluded) providers to score their probability of joining the OIG Exclusion List

- > Step 8.1:
 - Input your Scoring dataset our example used a sample of 100 non-excluded providers. The dataset included the 5 indicators and used the final weights to score their probability of being excluded. The final weights are inputted in columns K to P in our example.
- Step 8.2:
 - The formulas embedded in columns G to I will generate values based on the data inputted.

Note: the value of the mathematical constant "e" (= 2.7183, copied to the cell N3) is used to compute values in column G (labeled L).

The probabilities of exclusion P(X) of provider in row 2 are generated by using the formula "=H2/(1 + H2)", where the values of H2 are computed in the previous operation.