MiSUG, March 13, 2019
Transforming Ordinal Predictors to Numeric for Binary Logistic Models

Bruce Lund
Independent Consultant, Novi, MI

OneMagnify of Detroit, Manager of Analytics, Consultant
Ford Motor Company, Customer Database Manager
University of New Brunswick, Canada, Mathematics Professor
Education: Math PhD, Stanford University

USERS PROGRAM

SAS® GLOBAL FORUM 2019

Contact Me

Contact me for the paper I'll be giving at SGF (in April)

Also contact me for macro mentioned in this talk: %LOGIT_SCREEN_2

Blund_data@mi.rr.com … or … Blund.data@gmail.com

Predictors with Ordinal Levels

Examples

- Long-term bond credit ratings assigned by S&P:
  AAA, AA, A, BBB, BB, B, CCC, CC, C, D
- Do you approve or disapprove of way <President> is handling the job?
  Strongly or only moderately? (4 levels)
- Medical diagnosis stage: 0, 1, 2, 3, 4 (ordered categories)
- Medical treatment: No dose, low dose, medium dose, high dose
- Number of children in HH: 0, 1, 2, 3, 4+ (not quite numeric)

Two Questions about Ordinal X

- Does ordinal X have any predictive power? … look at IV (next slide)
- If so, how to use X?
  - Can use CLASS (dummies) or WOE (WOE: next slide)

    … But If X has K levels, then K-1 d.f. are being used

    OR
  - Somehow Recode Ordinal as numeric

    How and when should this be done ?

WOE transform and Information Value IV

X	Y = 0	Y = 1	Col % Y=0 "b_k"	Col % Y=1 "g_k"	Log(g_k/b_k) = X_woe	g_k - b_k	IV Terms = (g_k - b_k) *Log(g_k/b_k)
X1	2	1	25.0%	12.5%	-0.69315	-0.125	0.08664
X2	1	1	12.5%	12.5%	0.00000	0	0.00000
X3	5	6	62.5%	75.0%	0.18232	0.125	0.02279
SUM	8	8	100%	100%		IV =	0.10943

Siddiqi, *Intelligent Credit Scoring*

IV Range	Interpretation
IV < 0.02	"Not Predictive"
IV in [0.02 to 0.1)	"Weak"
IV in [0.1 to 0.3)	"Medium"
IV ≥ 0.3	"Strong"

X_woe *uses information* from Y … equal to K-1 d.f. Here, that is 2 d.f.

TestData has X1-X4 (all ORDINAL) and Target Y

DON'T look at Code !

```
DATA TestData;
Length X1-X4 $4;
Do I = 1 To 500;
/* N1 appears as Linear in Xbeta */
  If MOD(I,5) = 0 then N1 = 0;
   Else if MOD(I,5) = 1 then N1 = 2;
   Else if MOD(I,5) = 2 then N1 = 2.5;
   Else if MOD(I,5) = 3 then N1 = 3;
   Else if MOD(I,5) = 4 then N1 = 5;
  X1 = PUT(N1,Z3.1);
/* N2 appears as Log(N2) in Xbeta */
  N2 = (ranuni(1)> 0.5) + 1;
  If N2 > 1 then N2 = 2 + floor(10*(1-ranuni(1))**2);
  X2 = PUT(N2,Z2.);
/* N3 appears as (N3-3)**2 in Xbeta */
  N3 = Max(Floor(rannor(1) + 3.5),1);
  X3 = PUT(N3,Z1.);
Random1 = ranuni(1);
  If Random1 < 0.50 then X4 = "A1";
   Else If 0.50 <= Random1 < 0.55 then X4 = "A2";
   Else If 0.55 <= Random1 < 0.60 then X4 = "A3";
   Else If 0.60 <= Random1 < 0.70 then X4 = "B ";
   Else If 0.70 <= Random1 < 0.80 then X4 = "C ";
   Else If 0.80 <= Random1 < 0.90 then X4 = "D ";
   Else X4 = 'F ';
Xbeta = 1 + 3*rannor(1) + 0.5*N1 - 0.01*LOG(N2) -
        0.2*(N3-3)**2 - 1.0*(X4="A1") - 0.25*(X4="A2") +
        0.05*(X4="C") + 0.10*(X4="D") + 1.0*(X4="F");
    Y = (Exp(Xbeta)/(1 + Exp(Xbeta)) > 0.75);
    Output;
    End;
run;
```

Predictor	# Levels
X1	5
X2	11
X3	6
X4	7
Sum =	29

%LOGIT_SCREEN_2 (TestData, Y, , X1 X2 X3 X4);

VAR_NAME	Levels	IV (Info Value)	Siddiqi
X1	5	0.177	Med
X2	11	0.101	Med
X3	6	0.061	Weak
X4	7	0.172	Med

IV Range	Interpretation
IV < 0.02	"Not Predictive"
IV in [0.02 to 0.1)	"Weak"
IV in [0.1 to 0.3)	"Medium"
IV ≥ 0.3	"Strong"

All the Dummies

Overfit?

```
PROC LOGISTIC DATA = TestData;
CLASS X1 X2 X3 X4;
MODEL Y = X1 X2 X3 X4;
run;
```

Model c (=c-stat) is 0.690
-2*Log-Likelihood is 627.3

There are 25 parameters (excluding intercept).

… Can a good model be obtained if some of the CLASS effects are replaced with “LINEAR effects”?

All the WOE's

Degrees of freedom are unknown … somewhere between 4 and 25

```
PROC LOGISTIC DATA = TestData;
MODEL Y = X1_WOE  X2_WOE  X3_WOE  X4_WOE;
run;
```

Model c (=c-stat) is 0.685 … vs 0.690 (dummy)
-2*Log-Likelihood is 629.4 … vs 627.3 (dummy)
WOE is usually similar to Dummies (but ≤ fit)

There are 25 variable *recodings* (the WOE formulas)

… Can a good model be obtained if some of the WOE's are replaced with "LINEAR effects"?

"LINEAR effect" begs question of how to recode X

Simple Solution used today:

- Recode levels of X with consecutive integers starting at 1

  ➔ Levels for X: A, B, C, D ➔ Levels for X_n: 1, 2, 3, 4.

Does "recoded linear" give up *too much fit* vs. "dummies"

GOALS of following SLIDES:

- Give Test of Fit of recoded linear vs. dummies
- Give *Efficient Flexible* SAS to do Recoding and Testing

The Test of Fit: Recoded Linear vs Dummies

Does recoded linear X_n give up too much fit vs dummies?

- CLASS: PROC LOGISTIC; CLASS X; MODEL Y=X;
- RECODED: PROC LOGISTIC; MODEL Y=X_n; /* <= recoded*/

Compare chi-square statistics of model fit:

- $\chi^2_{\text{Compare}} = \chi^2_{\text{Class}} - \chi^2_{\text{Recoded Linear}}$ … k-2 d.f. if X has k levels
- H_0: Accept X_n vs. H_A Accept CLASS X (dummy)

  Suppose (using the data) that $\chi^2_{\text{Compare}} = C$

  If $P(\chi^2_{\text{Compare}} > C) < 0.05$ (or favorite α), then accept CLASS.

- If χ^2_{Compare} NOT Significant, then accept H_0 X_n

Code to Create X1_n, X2_n, X3_n, X4_n - no hard coding

Only need to modify what is in PURPLE … (good skeleton for MACRO)

- Step 1: PROC SUMMARY with CLASS X1-X4 with COMPLETETYPES. Output to Step1Out.
- Step 2: Recoding to linear. Output to Step2Out

Uses BY (Step1Out is sorted by X1-X4) and FIRST. and LAST. processing

- Step 3: PROC SUMMARY with CLASS X1-X4 Y; Output to Step3Out
- Step 4: MERGE Step2Out and Step3Out … BY X1-X4

Creates data set Step4Out - ready for use in PROC LOGISTIC

Next: MODEL COMPARISONS TEST … X_n versus CLASS X

MODEL COMPARISON TEST, Step 1

Get Chi-Squares for Test

```
ODS EXCLUDE ALL;
ODS OUTPUT EffectNotInModel = EffectNotInModel;
PROC LOGISTIC DATA = Step4Out;
CLASS  X1 - X4;  /* X1 - X4 are converted to dummies */
MODEL  Y = X1_n  X1   X2_n  X2   X3_n  X3   X4_n  X4  /* Recoded AND Class */
/ SELECTION = FORWARD  SLE = 1  STOP = 1  DETAILS;
FREQ _freq_;
run;
ODS EXCLUDE NONE;
PROC PRINT DATA = EffectNotInModel;
run;
```

(Score) Chi-Square of each predictor if entered as of the *First Step*

This is what is needed for MODEL COMPARISON TEST

MODEL COMPARISON TEST, Step 1

Display Chi-Squares for MODEL Fit

Obs	Effect	DF	Score ChiSq	ProbChiSq
1	X1_n	1	18.6151	<.0001
2	X1 (class)	4	21.5781	0.0002
3	X2_n	1	1.5375	~~0.2150~~
4	X2 (class)	10	12.1026	~~0.2782~~
5	X3_n	1	0.2256	0.6348
6	X3 (class)	5	7.5899	0.1803
7	X4_n	1	18.3148	<.0001
8	X4 (class)	6	20.2135	0.0025

PROC PRINT
DATA= EffectNotInModel;
run;

- DROP X2

X3 (class) “borderline”

MODEL COMPARISON TEST, Final

Requires a short, simple DATA Step

Is Recoded Linear good enough? $\chi^2_{Compare} = \chi^2_{Class} - \chi^2_{Recoded\ Linear}$

Var_Name	DF1	DF2	ChiSq1	ChiSq2	Compare Significance
X1_n	1	4	18.6151	21.5781	0.3974 => recode linear
~~X2_n~~	~~1~~	~~10~~	~~1.5375~~	~~12.1026~~	~~0.3067 => drop~~
X3_n	1	5	0.2256	7.5899	0.1178 => use CLASS
X4_n	1	6	18.3148	20.2135	0.8630 => recode linear

Fit versus Parsimony

Full Class Model vs. Reduced Model	d. f.	AIC	SBC	MODEL c
CLASS X1 - X4; MODEL Y= X1 - X4;	26	679.2	788.8	0.690
CLASS X3; MODEL Y = X1_n X3 X4_n;	8	660.7	694.4	0.669

Noticeable Loss of MODEL c

Large improvement in parsimony

Last Comment

Used Today: Simple recoding of Ordinal to Linear:

A ➔ 1, B ➔ 2, C ➔ 3, D➔ 4

SAS program could be enhanced to add other recodes:

e.g. Ordinal X recoded to LOG(X)

A ➔ log(1), B ➔ log(2), C ➔ log(3), D➔ log(4)

But, if OVER DONE, then this is data ransacking.
May as well use dummies.

Added Last Thought

Interpret coefficient of ordinal after recoding?

Predictor X: Poor, Fair, Good → X_n: 1, 2, 3

MODEL Y = X_n;

Not meaningful to say: β_X * (Fair - Good) = △ Log-Odds

Instead, discuss and compare these values:

Log-Odds at Fair (… X_n =1)
Log-Odds at Good (… X_n =2)

This discussion and comparison makes sense if MODEL Fits.

… A *semantic* distinction which avoids giving X an interval scale

Contact Me

Contact me for the paper I'll be giving at SGF (in April)

Also contact me for macro mentioned in this talk: %LOGIT_SCREEN_2

Blund_data@mi.rr.com … or … Blund.data@gmail.com