

Paper 22-2010

Adequacy and information values: Two tools for ranking predictors in generalized linear models

Doug Thompson, Assurant Health, Milwaukee, WI

ABSTRACT

Which predictors are most strongly associated with the predicted outcome? In business settings, this is one of the first questions that clients typically ask about predictive models. To answer this question, it is necessary to rank the predictors by some measure of strength of association with the predicted outcome. Adequacy and information values were the two ranking methods recommended in an earlier MWSUG paper (Thompson, 2009). This paper expands on the earlier one by presenting SAS macros and code that serve three purposes: 1) they rank predictors by Adequacy and information values in a variety of generalized linear models (logistic, poisson, etc.); 2) they rank predictors on Adequacy, taking into account continuous variables that are non-linearly associated with the outcome; and 3) they present a possible strategy for taking interactions into account in the rankings. The overall goal of this presentation is to provide predictive modelers with tools that they can use to more clearly explain model results to their clients.

INTRODUCTION

Generalized linear models are defined by three characteristics (Agresti, 2002; McCullagh & Nelder, 1999): 1) a *random component*, specifying a probability distribution for the response/outcome variable; 2) a *systematic component*, defining a weighted linear combination of the predictor variables; and 3) a *link function* relating the systematic and random components through a function such the logit (for logistic regression), log (Poisson regression) or identity (ordinary multiple regression). Frequently used subtypes of generalized linear models include:

- 1) Models for continuous outcomes (normal distribution, identity link function). Such models might be used to predict outcomes such as minutes of long-distance telephone usage or body mass index (BMI).
- 2) Models for binary outcomes, including logistic regression (binomial distribution, logit link function). Such models might be used to predict outcomes such as buy vs. no buy or insurance policy lapse vs. policy renewal.
- 3) Models for counts, including Poisson regression (Poisson distribution, log link) and negative binomial regression (negative binomial distribution, log link). Such models might be used to predict outcomes such as the count of property insurance claims due to a hurricane in a specific zip code, or the count of H1N1 ("swine flu") cases in a county.

Generalized linear models can be used to predict a wide variety of outcomes. There are other techniques could also be used to predict such outcomes, such as CART and neural networks, but it is

often difficult to outperform generalized linear models using these other techniques, at least not by a margin that is practically significant.

Generalized linear models employ a set of independent variables (“predictors”) to predict the outcome. A frequently asked question is, which variables in the predictor set were most strongly associated with the outcome? To answer this question, it is necessary to rank the predictors by some measure of association with the outcome. In ordinary least squares regression, predictors are often ranked by partial correlation. However, there is less consensus on how best to rank predictors in different subtypes of generalized linear models. Thompson (2009) described and compared six options for ranking predictors in one subtype of generalized linear model, logistic regression: 1) standardized coefficients, 2) p-values of Wald chi-square statistics, 3) a pseudo partial correlation metric for logistic regression, 4) Adequacy, 5) c-statistics, and 6) information values. That paper concluded that one of the most important considerations when selecting a ranking method is whether the method yields a useful interpretation that can be readily communicated to the target audience. For example, Adequacy has an interpretation that is relatively easy to understand and communicate – it is the proportion of the total explainable variation in the outcome (expressed as $-2LL$ in the model with all predictors) that one could explain if one had only a single predictor available; in other words, it is the explanatory value of a single predictor relative to the entire set of predictors. Ranking by information values is popular in some areas of analytics, and this is an excellent method for exploratory analyses where the predictors may have complex non-linear relationships with the predicted outcome. However, information values are not necessarily intuitive and they may be challenging to explain to some audiences. If the intuitiveness of Adequacy could be combined with the flexibility of information values, this might be an ideal method for ranking predictors in generalized linear models.

Definition of Adequacy and information values

Adequacy. Adequacy is the proportion of the full model log-likelihood that is explainable by each predictor individually (Harrell, 2001). One can think about this in terms of the explanatory value of each predictor individually, relative to the entire set. This equates “explanatory value” with $-2 \times \log$ likelihood ($-2LL$). The full model (with all predictors) provides the maximal $-2LL$. $-2LL$ can also be computed for each predictor individually (i.e., $-2LL$ in a model including only a single predictor) and compared with the full model $-2LL$. The greater the ratio, the more of the total $-2LL$ that could be explained if one had only the individual predictor available. An advantage of this metric is its interpretability – it indicates the proportion of the total explained variation in the outcome (expressed in terms of $-2LL$) that could be explained by a single predictor. Another strength is that Adequacy is entirely based on likelihood ratios, so it may be more reliable in small-to-medium samples, compared with methods that rely on Wald or score statistics (Agresti, 2002). A downside is that Adequacy can appear large, even if a predictor is weakly associated with the outcome – because Adequacy is relative to the predictive value of the entire set of predictors, if the full model has a small $-2LL$, an individual predictor can have a large Adequacy but little explanatory value for predicting the outcome. The best solution to this problem may be to only compute Adequacy for models with good predictive performance.

Information values. Information values represent the “total strength of the characteristic” (Siddiqi, 2006, p. 81) or predictor in a model. Information values are commonly used in data mining (including SAS Enterprise Miner) and marketing analytics. Information values provide a way of quantifying the amount of information about the outcome that one gains from a predictor. Larger information values indicate that a predictor is more informative about the outcome. One rule of thumb is that information values less than 0.02 indicate that a variable is not predictive; 0.02 to 0.1 indicate weak predictive power; 0.1 to 0.3 indicate medium predictive power; and 0.3+ indicates strong predictive power (Siddiqi, 2006). The information value method of ranking predictors is based on an analysis of each predictor in turn, without taking into account the other predictors. Information values have an important advantage when dealing with continuous predictors – to compute the information value, continuous predictors are broken into categories, and if there is a large difference within any category, the information value will be large. Thus information values are sensitive to differences in the outcome at any point along a continuous scale. There are various formulas for computing information values (e.g., Witten & Frank, 2005); this paper uses the definition described by Siddiqi (2006), as implemented in the %info_values macro provided below.

METHODS

To illustrate the methods, data from the National Health and Nutrition Examination Survey (NHANES), 2005-2006, was used. This survey was conducted by the National Center for Health Statistics, which is part of the Centers for Disease Control and Prevention. The data are publically available for free download (<http://www.cdc.gov/nchs/nhanes.htm>).

Three outcomes from NHANES 05-06 are used in the examples presented below:

- 1) *Drank beer* sometime during the past year (coded as 1=yes or 0=no, based on a food frequency questionnaire);
- 2) *Grams of beer consumed*, as reported in a single 24-hour dietary intake recall (the analysis was limited to those who reported drinking beer in the recall, that is, those consuming 0 grams of beer were excluded from the analysis; due to skewness of the distribution, grams of beer consumed was log transformed prior to analysis); and
- 3) *Number of days eating cheese* per week in the past year (based on a food frequency questionnaire).

The outcomes were modeled as a function of demographics, nutrient intake and eating behavior (e.g., eating vs. skipping breakfast). The predictors included a mix of continuous variables (age, energy intake, BMI) and binary variables (all others). The continuous variables differed in scale. Importantly, prior to analysis, observations with any missing data were excluded. The methods described below only yield valid rankings when used with complete data, otherwise ranking may be impacted by using different subsets of observations in different analyses. In practice, one would need to carefully consider the method for handling missing data as well as other analytic issues (e.g., whether or not linear

representation is appropriate for continuous variables, influence and residual diagnostics), but these are deemphasized in this paper because the sole purpose was to illustrate approaches for ranking predictors. For purposes of this illustration, the sample is treated as a simple random sample (survey weights and design variables are not used). In an analysis that is meant to be generalized to the population, weights and survey design would need to be taken into account in the analysis. After excluding the observations with missing data for the predictors and any participants under the age of 21, the sample included 4,064 observations.

Generalized linear models were used to model the outcomes. Logistic regression (binomial distribution, logit link function) was used to model drinking beer. Ordinary multiple regression (normal distribution, identity link function) was used to model log grams of beer consumed. Negative binomial regression (negative binomial distribution, log link function) was used to model days eating cheese (this was a better fit than Poisson regression, which would be a possible alternative method for count data).

Macros for estimating Adequacy

The macro %Adequacy_genmod_categ can be used to estimate Adequacy for generalized linear models. The input dataset is INSET and the dependent variable is DEPVAR. Predictors that will be modeled in their original scale without transformation (for example, binary predictors or continuous predictors for which a linear slope will be estimated) are listed in DO_NOT_CATEGORIZE. As illustrated below, it is convenient (but not necessary) to list such predictors in a space-delimited macro variable; below, this macro variable is called &do_not_categorize. Continuous predictors that will be split into categories, where each category will have a separate parameter estimate in the model (enabling modeling of non-linear associations between a given predictor and the outcome) are listed in CATEGORIZE. NUM_CATEGORIZE is the count of continuous predictors that will be split into categories prior to estimating the model. Such predictors are split into a set of ranked groups (the number of groups is NGROUPS, set to 10 as a default) using PROC RANK. PROC RANK creates ranked groups of approximately equal size, to the extent possible. The resulting categorical predictor created by PROC RANK is analyzed using a CLASS statement in PROC GENMOD. The NGROUPS default of 10 could be overridden, depending on the dataset. With very large datasets, breaking a predictor into more than 10 groups might enable modeling of more complex non-linear associations. Finally, LINK and DIST enable the user to specify the link function and response distribution of the model. The following sections of this paper present example macro calls and results for logistic regression (for the model of drinking beer, yes or no), multiple linear regression (for the model of log grams of beer consumed) and negative binomial regression (for the model of days eating cheese per week).

```
options mprint;
%macro
Adequacy_genmod_categ(inset=,depvar=,do_not_categorize=,categorize=,num_categ
orize=0,ngroups=10,link=,dist=);

* Execute loop if there are continuous covariates that will be modeled as
categorical;
%if &num_categorize ne 0 %then %do;
```

```

* For continuous variables that will be categorized, categorize based on
ranks.;
* The the ranks are grouped. The number of ranked groups is ngroups.;
proc rank data=&inset
out=ranked(drop=
%do ii=1 %to &num_categorize;
%let var1 = %scan(&categorize,&ii);
    &var1
%end;
)
groups=&ngroups;
var
%do jj=1 %to &num_categorize;
%let var2 = %scan(&categorize,&jj);
    &var2
%end;
;
ranks
%do kk=1 %to &num_categorize;
%let var3 = %scan(&categorize,&kk);
    rank_&var3
%end;
;
run;

* The result is a dataset where the covariates have the same name as;
* they originally had, but the original values are replaced by;
* a ranked grouping variable.;
data ranked(rename=(
%do ll=1 %to &num_categorize;
%let var3 = %scan(&categorize,&ll);
    rank_&var3 = &var3
%end;
));
set ranked;
run;

%end;

* If there are no continuous variables to be categorized, then make;
* no modifications to the dataset.;
%if &num_categorize = 0 %then %do;

data ranked;
set &inset;
run;

%end;

* Count the number of predictors that are not categorized;
proc contents data=ranked(keep=&do_not_categorize) out=names(keep=name)
noprint;
run;

```

```

data _null_;
set names end=eof;
retain num 0;
num+1;
if eof then call symput('numvar',num);
run;

* Compute log likelihood of model with all predictors;

ods noresults;
proc genmod data=ranked descending namelen=100;
%if &num_categorize ne 0 %then %do;
    class &categorize;
%end;
model &depvar = &categorize &do_not_categorize / link=&link dist=&dist;
ods output Modelfit=fit;
run;
ods results;

data fit2;
set fit;
interceptandcovariates=-2*value;
keep interceptandcovariates;
if Criterion='Log Likelihood';
run;

ods noresults;
proc genmod data=ranked descending namelen=100;
model &depvar = / link=&link dist=&dist;
ods output Modelfit=fit;
run;
ods results;

data fit3;
set fit;
interceptonly=-2*value;
keep interceptonly;
if Criterion='Log Likelihood';
run;

data allfit;
merge fit2 fit3;
run;

data _fit_all(rename=(diff=all_lr));
attrib var length=$100.;
set allfit;
all=1;
diff=interceptonly-interceptandcovariates;
keep all diff;
run;

* 2. Compute log likelihood for models with one predictor only;
* 2A. Compute for predictors that are not categorized.;
    %do j=1 %to &numvar;

```

```

%let curr_var = %scan(&do_not_categorize,&j);

ods noresults;
proc genmod data=ranked descending namelen=100;
model &depvar = &curr_var / link=&link dist=&dist;
ods output Modelfit=fit;
run;
ods results;

data fit2;
set fit;
interceptandcovariates=-2*value;
keep interceptandcovariates;
if Criterion='Log Likelihood';
run;

ods noresults;
proc genmod data=ranked descending namelen=100;
model &depvar = / link=&link dist=&dist;
ods output Modelfit=fit;
run;
ods results;

data fit3;
set fit;
interceptonly=-2*value;
keep interceptonly;
if Criterion='Log Likelihood';
run;

data allfit;
merge fit2 fit3;
run;

data _fit;
attrib var length=$100.;
set allfit;
var="&curr_var";
diff=interceptonly-interceptandcovariates;
keep var diff;
run;

%if &j=1 %then %do;
    data redfit;
    set _fit;
    run;
%end;
%else %do;
    data redfit;
    set redfit _fit;
    all=1;
    run;
%end;
%end;

```

```

* 2B. Compute for predictors that are categorized.;
%do k=1 %to &num_categorize;
%let curr_var = %scan(&categorize,&k);

ods noresults;
proc genmod data=ranked descending namelen=100;
class &curr_var;
model &depvar = &curr_var / link=&link dist=&dist;
ods output Modelfit=fit;
run;
ods results;

data fit2;
set fit;
interceptandcovariates=-2*value;
keep interceptandcovariates;
if Criterion='Log Likelihood';
run;

ods noresults;
proc genmod data=ranked descending namelen=100;
model &depvar = / link=&link dist=&dist;
ods output Modelfit=fit;
run;
ods results;

data fit3;
set fit;
interceptonly=-2*value;
keep interceptonly;
if Criterion='Log Likelihood';
run;

data allfit;
merge fit2 fit3;
run;

data _fit;
attrib var length=$100.;
set allfit;
var="&curr_var";
diff=interceptonly-interceptandcovariates;
keep var diff;
run;

%if &k=1 %then %do;
data redfit2;
set _fit;
run;
%end;
%else %do;
data redfit2;
set redfit2 _fit;
all=1;
run;

```



```

    %end;
    %end;

%if &num_categorize ne 0 %then %do;
data redfit;
set redfit redfit2;
run;
%end;

* 3. Compute the ratio of log likelihoods;

proc sql;
create table _redfit as select * from
redfit a, _fit_all b
where a.all=b.all;
quit;

data _redfit(rename=(var=variable));
set _redfit;
a_statistic=(diff/all_lr);
drop all;
run;

proc sort data=_redfit;
by descending a_statistic;
run;

%mend Adequacy_genmod_categ;

```

Below are example calls of the %Adequacy_genmod_categ macro. Results of these macro calls are described in the Results section.

```

* All predictors - linear slopes are estimated for continuous predictors;
%let vars=breakfast
female
mexam
othhisp
white
afam
inc_lt25K
inc_75Kplus
college_grad
some_college
hs_grad
married
RIDAGEYR
pregnant
energy
bmi;

* Binary predictors;
%let do_not_categorize=
breakfast
female
mexam

```

```

othhisp
white
afam
inc_lt25K
inc_75Kplus
college_grad
some_college
hs_grad
married
pregnant;

* Continuous predictors to categorize;
%let categorize=
energy
bmi
RIDAGEYR;

* For the analyses below, limit to obs where the outcome is not missing;
* (all observations with missing data on the predictors were already deleted;
* when the dataset was created).;
data nhanes_0506_drunkbeer;
set mwsug10.nhanes_0506_mwsug10(where=(drank_beer ne .));
run;

***** MODEL OF DRINKING BEER (1=YES, 0=NO);
* Adequacy with continuous predictors categorized;
%Adequacy_genmod_categ(inset=nhanes_0506_drunkbeer,
depvar=drank_beer,
do_not_categorize=&do_not_categorize,categorize=&categorize,num_categorize=3,
ngroups=10,link=logit,dist=bin);

* Adequacy with continuous predictors treated as continuous;
%Adequacy_genmod_categ(inset=nhanes_0506_drunkbeer,
depvar=drank_beer,
do_not_categorize=&vars,categorize=,num_categorize=0,ngroups=10,link=logit,dist=bin);

data nhanes_0506_gramsbeer;
set mwsug10.nhanes_0506_mwsug10(where=(ln_gm_beer ne .));
run;

***** MODEL OF GRAMS OF BEER CONSUMED;
* Adequacy in the model of grams of beer consumed, identity link and;
* normal distribution;
%Adequacy_genmod_categ(inset=nhanes_0506_gramsbeer,
depvar=ln_gm_beer,
do_not_categorize=&do_not_categorize,categorize=&categorize,num_categorize=3,
ngroups=10,link=identity,dist=normal);

data nhanes_0506_cheeseweek;
set nhanes_0506_mwsug10(where=(days_cheese_week ne .));
run;

***** MODEL OF DAYS EATING CHEESE PER WEEK;
* Adequacy in the model of days eating cheese per week, log link and;

```

```

* negative binomial distribution;
%Adequacy_genmod_categ(inset=nhanes_0506_cheeseweek,
depvar=days_cheese_week,
do_not_categorize=&do_not_categorize,categorize=&categorize,num_categorize=3,
ngroups=10,link=log,dist=negbin);

```

The following code was used to compute information values for the logistic regression model of drinking beer. With relatively minor modifications, this macro could also be used for other outcomes. The model is based on methods described by Siddiqi (2006).

```

* Information values, for comparison;
options mprint;
%macro info_values;
%let vars =
RIDAGEYR
bmi
energy
breakfast
female
mexam
othhisp
white
afam
inc_lt25K
inc_75Kplus
college_grad
some_college
hs_grad
married
pregnant;

%let binary =
0
0
0
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1;

%do i=1 %to 16;
%let curr_var = %scan(&vars,&i);
%let curr_binary = %scan(&binary,&i);

%if &curr_binary=0 %then %do;
proc rank data=nhanes_0506_drunkbeer out=iv&i groups=10 ties=high;

```

```

var &curr_var;
ranks rnk_&curr_var;
run;
%end;
%if &curr_binary=1 %then %do;
data iv&i;
set nhanes_0506_drankbeer;
rnk_&curr_var=&curr_var;
run;
%end;

proc freq data=iv&i(where=(drank_beer=1)) noprint;
tables rnk_&curr_var / out=out1_&i(keep=rnk_&curr_var percent
rename=(percent=percent1));
run;

proc freq data=iv&i(where=(drank_beer=0)) noprint;
tables rnk_&curr_var / out=out0_&i(keep=rnk_&curr_var percent
rename=(percent=percent0));
run;

data out01_&i;
merge out0_&i out1_&i;
by rnk_&curr_var;
_percent0=percent0/100;
_percent1=percent1/100;
sub_iv=(percent0-percent1)*log(percent0/percent1);
run;

proc means data=out01_&i sum noprint;
var sub_iv;
output out=sum&i sum=infoval;
run;

data _sum&i;
length variable $100;
set sum&i;
variable="&curr_var";
drop _type_ _freq_;
run;

proc datasets library=work nolist;
delete iv&i out1_&i out0_&i out01_&i sum&i;
run;
quit;

%if &i=1 %then %do;
data iv_set;
set _sum&i;
run;
%end;
%else %do;
data iv_set;
set iv_set _sum&i;
run;

```

```

%end;
%end;
%mend info_values;
%info_values;

```

RESULTS

Descriptive statistics for variables used in the analyses are shown in Table 1. There were 4,064 observations with complete data for the predictors. Each model was limited to observations with complete data for the outcome variable (n=2,757 for the model of drinking beer, n=567 for the model of log grams of beer consumed and n=2,934 for the model of days eating cheese per week).

Table 1. Descriptive statistics for variables used in the analyses.

Analysis variables	Mean / %	n
<i>Outcomes</i>		
Drank beer, last year	52%	2,757
Gm beer consumed *	1283.8	567
Days eating cheese / week	2.7	2,934
<i>Predictors</i>		
Ate breakfast *	84%	4,064
Daily energy intake *	2232.8	4,064
Female	52%	4,064
Mexican American	8%	4,064
Other Hispanic	3%	4,064
White	72%	4,064
African American	11%	4,064
Household income <\$25K	19%	4,064
Household income \$75K+	32%	4,064
College graduate	28%	4,064
Some college	31%	4,064
High school graduate	25%	4,064
Married	59%	4,064
Age (years)	46	4,064
Pregnant	2%	4,064
Body mass index (BMI)	28.6	4,064

* Based on 24-hour dietary intake recall

Adequacy (with continuous predictors categorized vs. not) and information values

Table 2 shows Adequacy and information values for predictors in the logistic regression model of drinking beer. Adequacy estimates are shown separately for analyses in which continuous predictors were split into categories prior to modeling (10 ranked groups were created and the categorized

predictors were analyzed using a CLASS statement in PROC GENMOD) and for analyses in which a linear slope was estimated for the continuous predictors.

Table 2. Adequacy and information values for predictors in the logistic regression model of drinking beer.

Predictors	Ranking methods		
	Adequacy: Linear slope for continuous variables	Adequacy: Decile classification of continuous variables	Information values
Female	0.520	0.486	0.326
Daily energy intake *	0.239	0.258	0.174
Age (years) *	0.096	0.149	0.123
African American	0.081	0.075	0.050
Household income <\$25K	0.065	0.061	0.041
White	0.064	0.060	0.040
Household income \$75K+	0.062	0.058	0.038
College graduate	0.059	0.055	0.037
BMI *	0.057	0.083	0.056
Pregnant	0.041	0.039	0.026
Ate breakfast	0.005	0.005	0.003
High school graduate	0.001	0.001	0.001
Some college	0.001	0.001	<0.001
Married	0.001	<0.001	<0.001
Other Hispanic	<0.001	<0.001	<0.001
Mexican American	<0.001	<0.001	<0.001

* Continuous predictors

The predictors in Table 2 are sorted in terms of Adequacy for the model in which linear slopes were estimated for continuous predictors. The ranking is the same for information values, except that BMI is ranked 9th for Adequacy with linear slopes estimated for continuous predictors, but 4th in terms of information values. In contrast, Adequacy with categorized continuous predictors had exactly the same rankings as the information value rankings – BMI is 4th by both methods. Both Adequacy with categorized continuous predictors and information values pointed to similar effect sizes for the predictors.

The two methods for computing Adequacy yielded markedly different Adequacy estimates for some of the continuous predictors, particularly age and BMI. For both of these predictors, Adequacy was higher when they were categorized prior to analysis. This is because age and BMI were non-linearly associated with drinking beer, as shown in Figures 1 (BMI) and 2 (age) below. This explains why Adequacy is greater in models that enable more complex representations than a linear slope.

Figure 1. Percent of individuals who drank beer by BMI.

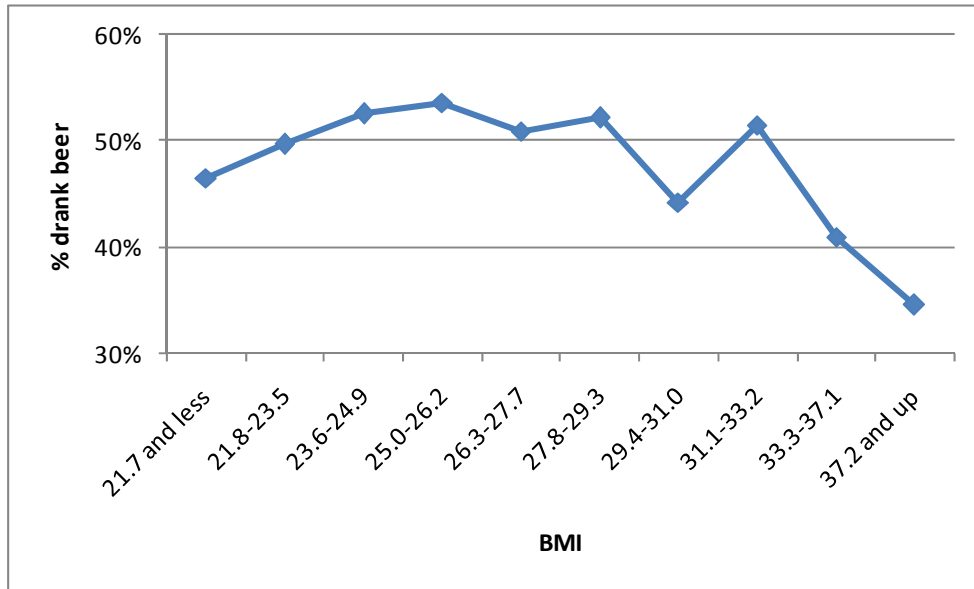
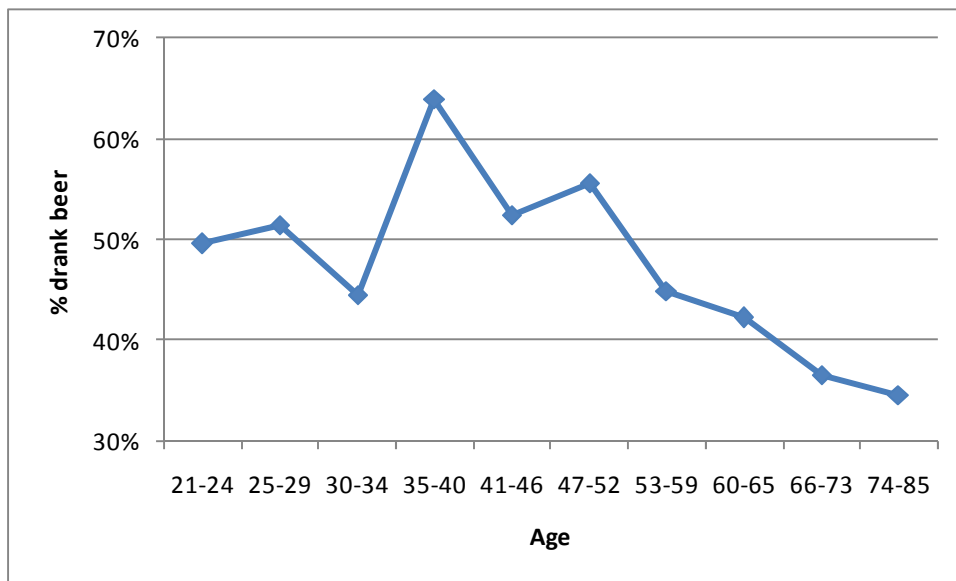


Figure 2. Percent of individuals who drank beer by age.



Adequacy for different types of generalized linear models

The %Adequacy_genmod_categ macro was also used to rank predictors in the models of log grams of beer consumed and days eating cheese per week. For both models, Adequacy was computed using the method where continuous predictors were split into categories and analyzed using a CLASS statement in PROC GENMOD. Adequacy for predictors in the model of log grams of beer consumed had some

similarity with Adequacy for predictors in the model of drinking beer (e.g., gender, energy intake, age and BMI all had high Adequacy), except that energy intake had much higher Adequacy in the model of log grams of beer consumed (this makes sense – drinking a large amount of beer causes high caloric intake). It also makes sense that the Adequacy of predictors in the model of eating cheese would be different from Adequacy in the beer models – there is no reason to expect that drinking beer and eating cheese would have similar correlates (for example, some teetotalers eat cheese).

Table 3. Adequacy for predictors in the models of log grams of beer consumed and days eating cheese per week.

Predictors	Model	
	Grams beer consumed, 24-hr period	Days eating cheese per week
Female	0.1886	<0.0001
Daily energy intake *	0.5122	0.1810
Age (years) *	0.1137	0.2728
BMI *	0.1014	0.0437
African American	0.0035	0.0941
Household income <\$25K	0.0044	0.0553
White	0.0013	0.2943
Household income \$75K+	0.0160	0.1006
College graduate	0.0637	0.1706
Pregnant	0.0173	0.0302
Ate breakfast	0.0413	0.0002
High school graduate	0.0008	0.0293
Some college	0.0002	0.0366
Married	0.0173	0.0047
Other Hispanic	0.0343	<0.0001
Mexican American	0.0201	0.0912

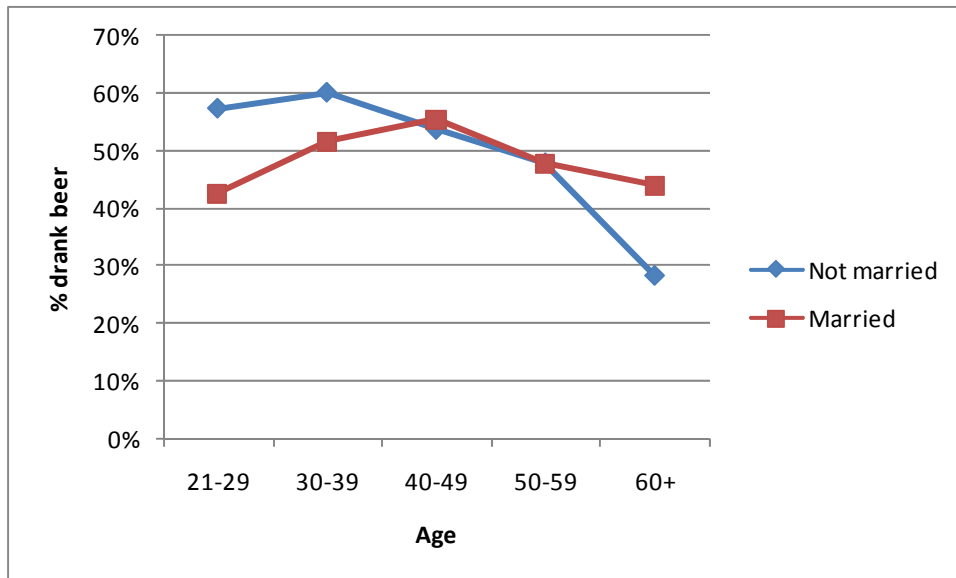
* Continuous predictors

Computing Adequacy when there are important interactions among the predictors

All of the analyses considered so far have looked at main effects only; predictors were ranked in terms of main effects and possible interactions were ignored. However, there may be important interactions among the predictors in some models. In this section, possible strategies for estimating Adequacy in models with interactions are discussed.

As an example, in the model of drinking beer, there is evidence of an interaction between age and marriage status. The interaction is graphed below (Figure 3).

Figure 3. Percent of individuals who drank beer by marriage status and age.



Although the age effect looks quadratic, we will estimate a linear effect for purposes of this example, to keep the illustration simple.

The interaction is significant in the following model estimated via PROC GENMOD:

Model with interaction:

```
proc genmod data=nhanes_0506_mwsug10 descending;
model drank_beer =
married
RIDAGEYR
married*RIDAGEYR / dist=bin link=logit;
run;
```

Main effects only model:

```
proc genmod data=nhanes_0506_mwsug10 descending;
model drank_beer =
married
RIDAGEYR / dist=bin link=logit;
run;
```

The model equations estimated via PROC GENMOD are:

Model with interaction:

$$\text{Log odds of drinking beer} = 1.1614 - 1.1142 * \text{married} - 0.0272 * \text{age} + 0.0247 * \text{married} * \text{age}$$

Model with main effects only:

$$\text{Log odds of drinking beer} = 0.5391 + 0.0659 * \text{married} - 0.0139 * \text{age}$$

Predictor values can be plugged into these models and predictions obtained. Results are shown in Table 4. The predictions were transformed from the log odds scale back to the probability scale, which yields the predicted percentages.

Table 4. Predicted percent of individuals drinking beer for selected values of marriage status and age, in models with vs. without the marriage by age interaction.

Predictor values	Model with interactions		Model without interactions	
	Predicted %	Difference (married - not married)	Predicted %	Difference (married - not married)
Age 30, married	49%		51%	
Age 30, not married	59%	-9%	53%	-2%
Age 50, married	48%		44%	
Age 50, not married	45%	3%	46%	-2%

The interaction indicates that the difference between married and unmarried individuals varies by age. Specifically, married 30-year-olds are predicted to drink beer at lower rates than unmarried 30-year-olds, while there is a slight trend for married 50-year-olds to drink beer at greater rates than unmarried 50-year-olds. In contrast, for the model without the interaction, unmarried individuals are predicted to drink beer at lower rates than married individuals, regardless of age.

Essentially, the interaction modifies the predicted marriage effect, conditional on age. Conversely, one could view the interaction as modifying the predicted age effect, conditional on marriage status. The total explanatory value of marriage and age are masked if we ignore the interaction. The age by marriage status interaction does have important explanatory value. The model without the interaction fits the data less well (BIC = 3774 vs. 3798 for the models with vs. without the interaction, respectively) and does not capture the age and marriage status trends that are apparent Figure 3.

There are different ways that one could deal with the interaction. One strategy would be to include the interaction term as if it were a separate predictor, then estimate Adequacy as if the interaction term was just another predictor. However, comparing such interaction terms with main effects would be counter-intuitive and possibly confusing. An alternative approach would be to say that for purposes of computing Adequacy, when the model includes main effects for A and B as well as the A x B interaction, then the “total Adequacy” of A is the Adequacy of A *plus* A x B, while the while the total Adequacy of B is the Adequacy of B *plus* A x B. If the interaction is small or non-existent, then “main effects Adequacy” will be close to total Adequacy. The larger the interaction, the greater the difference between total Adequacy and main effects Adequacy.

The total Adequacy approach to dealing with interactions is illustrated in the model of drinking beer as follows. The code shows the models needed to compute Adequacy in this situation. For results, only the log likelihoods are shown, because this is all that is needed for computing Adequacy.

```

* Example of estimating Adequacy;
ods listing;
proc genmod data=nhanes_0506_mwsug10 descending;
model drank_beer = / dist=bin link=logit;
run;
* Log Likelihood -1907.9877;
* -2*-1907.9877 = 3815.9754 (1);
ods listing;
proc genmod data=nhanes_0506_mwsug10 descending;
model drank_beer =
married
RIDAGEYR
married*RIDAGEYR / dist=bin link=logit;
run;
* Log Likelihood -1871.3846;
* -2*-1871.3846 = 3742.7692 (2);
ods listing;
proc genmod data=nhanes_0506_mwsug10 descending;
model drank_beer =
married
married*RIDAGEYR / dist=bin link=logit;
run;
* Log Likelihood -1907.5045;
* -2*-1907.5045 = 3815.009 (3);
ods listing;
proc genmod data=nhanes_0506_mwsug10 descending;
model drank_beer =
RIDAGEYR
married*RIDAGEYR / dist=bin link=logit;
run;
* Log Likelihood -1883.6946;
* -2*-1883.6946 = 3767.3892 (4);

* Main effects models for comparison;
ods listing;
proc genmod data=nhanes_0506_mwsug10 descending;
model drank_beer =
married / dist=bin link=logit;
run;
* Log Likelihood -1907.8766;
* -2*-1907.8766 = 3815.7532 (5);
ods listing;
proc genmod data=nhanes_0506_mwsug10 descending;
model drank_beer =
RIDAGEYR / dist=bin link=logit;
run;
* Log Likelihood -1887.4815;
* -2*-1887.4815 = 3774.963 (6);

```

Adequacy of marriage status:

$$-2LL, \text{ full model} = 3815.9754 (1) - 3742.7692 (2) = 73.2062$$

$$-2LL, \text{ model with marriage and marriage} \times \text{age} = 3815.9754 (1) - 3815.009 (3) = 0.9664$$

Total Adequacy of marriage status = $0.9664 / 73.2062 = 0.013201$

Compared to main effects Adequacy of marriage status:

-2LL, model with marriage status = $3815.9754 (1) - 3815.7532 (5) = 0.2222$

Main effects Adequacy of marriage status = $0.2222 / 73.2062 = 0.003035$

Adequacy of age:

-2LL, full model = $3815.9754 (1) - 3742.7692 (2) = 73.2062$

-2LL, model with age and marriage x age = $3815.9754 (1) - 3767.3892 (4) = 48.5862$

Total Adequacy of age = $48.5862 / 73.2062 = 0.66369$

Compared to main effects Adequacy of age:

-2LL, model with age = $3815.9754 (1) - 3774.963 (6) = 41.0124$

Main effects Adequacy of age = $41.0124 / 73.2062 = 0.560231$

Using this method to estimate Adequacy in a model with interactions, the total Adequacy of age and marriage status are greater than the main effects Adequacy of these variables, which is what we would expect when an important interaction is present.

The tentative recommendation is to use the total Adequacy approach to compute Adequacy when there are important interactions in a model. There may also be other approaches to estimate Adequacy in the presence of interactions; these await further research.

CONCLUSION

Adequacy is a useful tool for ranking predictors in terms of their explanatory value in generalized linear models. For continuous predictors, Adequacy estimates can vary greatly depending on whether continuous predictors are categorized prior to analysis and modeled using a CLASS statement in PROC GENMOD (which enables modeling of non-linear patterns of association), as opposed to estimating linear slopes. The difference is a function of the linearity (or lack thereof) of association between the predictor and the outcome. In the examples presented in this paper, when non-linear associations were taken into account, Adequacy and information values yielded similar predictor rankings. Adequacy is a useful method for ranking predictors in a wide variety of generalized linear models, as illustrated in this paper. When there are important interactions in a model, it may make sense to estimate “total Adequacy” as opposed to “main effects Adequacy” (in the model $OUTCOME = A + B + AxB$, the total Adequacy of A is the Adequacy of A *and* AxB, while the total Adequacy of B is the Adequacy of B *and* AxB).

REFERENCES

Agresti A. (2002). Categorical data analysis. Hoboken, NJ: Wiley.

Harrell F. (2001). Regression modeling strategies. New York: Springer.

McCullagh P., Nelder JA. (1999). Generalized linear models. New York: Chapman & Hall/CRC.

Siddiqi N. (2006). Credit risk scorecards: Developing and implementing intelligent credit scoring. Hoboken, NJ: Wiley.

Thompson D. (2009). Ranking predictors in logistic regression. Proceedings of the 20th annual Midwest SAS User Group, Cleveland, OH (www.mwsug.org/proceedings/2009/stats/MWSUG-2009-D10.pdf).

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Doug Thompson

Assurant Health

500 West Michigan

Milwaukee, WI 53203

Work phone: (414) 299-7998

E-mail: Doug.Thompson@Assurant.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.