

---

# 27

---

## *Decile Analysis: Perspective and Performance*

---

---

### **Appendix 27.A Incremental Gain in Accuracy: Model versus Chance**

---

```
libname da 'c://0-da';
data dec;
  set da.score;
  PREDICTED=0;
  if prob_hat > 0.222 then PREDICTED=1;
run;

data dec_;
  set dec end=last;
  wght=1; output;
  if last then do;
    predicted=0; wght=0; output;
    predicted=1; wght=0; output;
  end;
run;

PROC FREQ data=dec_;
  table RESPONSE*PREDICTED / norow nocol sparse out=D;
  weight wght / zeros;
run;
```

```

PROC TRANSPOSE data=D out=transp;
run;

data COUNT;
set transp;
if _NAME_="COUNT";
array col(4) col1-col4;
array cell(4) a b c d;
do i=1 to 4;
cell(i)=col(i);
drop i col1-col4;
end;
a_d=a+d;
N=a+b+c+d;
if _NAME_="COUNT" then MODEL_ACCURACY= (a+d)/(a+b+c+d);
if _NAME_="COUNT" then MODEL_GEOM_ACCURACY= SQRT( (d/(c+d)) * (d/(b+d)));

drop a--d _NAME_ _LABEL_;
m=1;
run;

data PERCENT;
set transp;
if _NAME_="PERCENT";
array col(4) col1-col4;
array cell(4) a b c d;
do i=1 to 4;
cell(i)=col(i)/100;
drop i col1-col4;
end;
ab_sq=(a+b)**2;
cd_sq=(c+d)**2;
if _NAME_="PERCENT" then CHANCE_ACCURACY= (((a+b)**2)+((c+d)**2));
drop a--d _NAME_ _LABEL_;
m=1;
run;

```

```

PROC SORT data=COUNT; by m;
PROC SORT data=PERCENT; by m;
run;

data PERCOUNT;
merge PERCENT COUNT; by m;
drop m;
keep MODEL_ACCURACY CHANCE_ACCURACY MODEL_GEOM_ACCURACY;
run;

data ACCURACY;
set PERCOUNT;
ACCURACY_INCREMENTAL_GAIN=
  ((MODEL_ACCURACY - CHANCE_ACCURACY)/CHANCE_ACCURACY);
run;

PROC PRINT data=ACCURACY;
var MODEL_ACCURACY CHANCE_ACCURACY
    ACCURACY_INCREMENTAL_GAIN;
format CHANCE_ACCURACY MODEL_ACCURACY
    ACCURACY_INCREMENTAL_GAIN percent8.2;
run;

```

---

## Appendix 27.B Incremental Gain in Precision: Model versus Chance

---

```

libname da 'c://0-da';
options pageno=1;

data dec;
set da.score;
PREDICTED=0;
if prob_hat > 0.222 then PREDICTED=1;
run;

```

```

data dec_;
set dec end=last;
wght=1; output;
if last then do;
predicted=0; wght=0; output;
predicted=1; wght=0; output;
end;
run;

PROC FREQ data=dec_;
table RESPONSE*PREDICTED / norow nocol sparse out=D;
weight wght / zeros;
run;

PROC TRANSPOSE data=D out=transp;
run;

data Precision_IMPROV;
retain d;
set transp;
array col(4) col1-col4;
array cell(4) a b c d;
do i=1 to 4;
cell(i)=col(i);
drop i col1-col4;
end;

b_d=b+d;
if _NAME_="COUNT" ;
if _NAME_="COUNT" then MODEL_Precision= d/(b+d);
c_d=c+d;
N=a+b+c+d;
if _NAME_="COUNT" then CHANCE_Precision= (c+d)/N;
PRECISION_INCREMENTAL_GAIN=MODEL_Precision - CHANCE_Precision;
drop a--c _NAME_ _LABEL_;
run;

```

```
PROC PRINT data=Precision_IMPROV;
var MODEL_Precision CHANCE_Precision PRECISION_INCREMENTAL_GAIN;
format MODEL_Precision CHANCE_Precision PRECISION_INCREMENTAL_GAIN
percent8.2;
run;
```

---

## **Appendix 27.C RESPONSE Model Decile PROB\_est Values**

---

```
libname ca 'c:\0-PCA_CCA';
options pageno=1;
title' completes X10-X14';
```

```
PROC LOGISTIC data=ca.completes nosimple des noprint outest=coef;
model RESPONSE = X10-X14;
run;
```

```
PROC SCORE data=ca.completes predict type=parms score=coef out=score;
var X10-X14;
run;
```

```
data score;
set score;
estimate=response2;
label estimate='estimate';
run;
```

```
data notdot;
set score;
if estimate ne .;
```

```
PROC MEANS data=notdot noprint sum; var wt;
output out=samsize (keep=samsize) sum=samsize;
run;
```

```
data scoresam (drop=samsize);
set samsize score;
retain n;
if _n_=1 then n=samsize;
if _n_=1 then delete;
run;
```

```
PROC SORT data=scoresam; by descending estimate; run;
```

```
PROC FORMAT;
value Decile
0='top '
1='2'
2='3'
3='4'
4='5'
5='6'
6='7'
7='8'
8='9'
9='bottom';
```

```
data complete_probs;
set scoresam;
if estimate ne . then cum_n+wt;
if estimate = . then Decile=.;
else Decile=floor(cum_n*10/(n+1));
PROB_est=exp(estimate)/(1+ exp(estimate));
keep PROB_est response wt Decile; run;
```

```
PROC TABULATE data=complete_probs missing;
class Decile;
var response PROB_est;
table Decile all, (n*f=comma8.0 response *(mean*f=8.4 sum*f=comma8.0)
(PROB_est) *(( min max)*f=8.4));
format Decile Decile.;
run;
```

---

## Appendix 27.D 2 X 2 Tables by Decile

---

```
libname da 'c://0-da';
%let problist=%str(0.4602, 0.2654, 0.2354, 0.2138, 0.2008, 0.1782, 0.1427, 0.1060, 0.0682,
0.02228);
%macro dochance;
  %do k=0 %to 9;
    %let count=%eval(&k+1);
    %let prob=%qscan(&problist,&count,%str(', '));
    %put k=&k prob=&prob;

    title2 "Decile=&k PROB_est=&prob ";
    data dec&k;
      set da.score;
      if dec=&k;
        ACTUAL=RESPONSE;
        PREDICTED=0;
        if prob_hat > &prob then PREDICTED=1;
    run;

    data dec_&k;
      set dec&k end=last;
      wght=1; output;
      if last then do;
        predicted=0; wght=0; output;
        predicted=1; wght=0; output;
      end;
    run;

PROC FREQ data=dec_&k ;
table ACTUAL*PREDICTED / norow nocol sparse out=D&k;
weight wght / zeros;
run;
```

```
PROC TRANSPOSE data=D&k out=transp&k;  
run;
```

```
%end;  
%mend;  
%doChance
```

```
title' '  
data PERCENT;  
retain Decile;  
set transp0 transp1 transp2 transp3 transp4  
transp5 transp6 transp7 transp8 transp9;  
if _NAME_="PERCENT" ;  
if _NAME_="PERCENT" then CHANCE_Decile_Precision= (col4/(col2+col4));  
drop _LABEL_ _NAME_;  
Decile+1;  
array col(4) col1-col4;  
array cell(4) a b c d;  
do i=1 to 4;  
cell(i)=col(i);  
drop i ;  
end;  
run;
```

```
PROC PRINT data=PERCENT noobs;  
var Decile CHANCE_Decile_Precision a b c d;  
title'Chance Decile Precision';  
format a b c d 5.2;  
format CHANCE_Decile_Precision 6.4;  
run;
```

```
data CHANCE_Decile_Precision;  
set PERCENT;  
keep CHANCE_Decile_Precision;  
PROC TRANSPOSE out=transCHANCE_Decile_Precision;  
var CHANCE_Decile_Precision;  
run;
```



```
PROC PRINT data=transCHANCE_Decile_Precision;
title 'data=transCHANCE_Decile_Precision';
run;
```

```
data CHANCE_Decile_Precision;
set transCHANCE_Decile_Precision;
CHANCE_Decile_Precision=
    geommean(col1, col2, col3, col4, col5, col6, col7, col8, col9, col10);
m=1;
keep CHANCE_Decile_Precision;
run;
```

```
PROC PRINT data=CHANCE_Decile_Precision noobs;
title' geometric mean - CHANCE_Decile_Precision ' ;
format CHANCE_Decile_Precision 6.4;
run;
```

```
title' ';
data COUNT;
retain Decile;
set
transp0 transp1 transp2 transp3 transp4
transp5 transp6 transp7 transp8 transp9;
if _NAME_="COUNT" ;
drop _LABEL_ _NAME_;
Decile+1;
array col(4) col1-col4;
array cell(4) a b c d;
do i=1 to 4;
cell(i)=col(i);
drop i ;
end;

N=a+b+c+d;
run;
```

```
PROC PRINT data=COUNT;
var Decile a b c d N;
title 'COUNT';
run;
```

```
PROC SUMMARY data=COUNT;
var col1 col2 col3 col4;
output out=sum_counts sum=;
run;
```

```
data sum_counts;
set sum_counts;
_name_="COUNT";
drop _TYPE_ _FREQ_;
sum_counts=sum(of col1-col4);
N=sum(of col1-col4);
MODEL_Decile_Precision=col4 / ( col2 + col4);
array col(4) col1-col4;
array cell(4) a b c d;
do i=1 to 4;
cell(i)=col(i);
drop i;
end;
m=1;
run;
```

```
PROC PRINT data=sum_counts; var MODEL_Decile_Precision a b c d N ;
format MODEL_Decile_Precision 6.4;
title 'sum_counts';
run;
```

```
PROC SORT data=CHANCE_Decile_Precision; by m;
PROC SORT data=sum_counts; by m;
run;
```

```
data PERCOUNTS;
merge CHANCE_Decile_Precision sum_counts;
run;
```

```
PROC PRINT data=PERCOUNTS;  
title 'PERCOUNTS';  
run;
```

```
data Decile_Precision_Gain;  
set PERCOUNTS;  
DECILE_PRECISION_INCREMENT_GAIN=  
((MODEL_Decile_Precision - CHANCE_Decile_Precision)/CHANCE_Decile_Precision);  
run;
```

```
title '';  
PROC PRINT data=Decile_Precision_Gain noobs;  
var MODEL_Decile_Precision CHANCE_Decile_Precision  
    DECILE_PRECISION_INCREMENT_GAIN;  
format MODEL_Decile_Precision CHANCE_Decile_Precision  
    DECILE_PRECISION_INCREMENT_GAIN percent8.2;  
title 'Decile_Precision_Gain';  
run;
```