

Statistical Presentations Using the Power of PROC TABULATE

C. Olivia Rud, VP of Analytic Services, DirectCom, West Chester, PA

ABSTRACT

PROC TABULATE is a versatile tool which can combine the functionality of PROC MEANS and PROC FREQ. Additional power exists in its ability to perform analysis on grouped data. This paper details the steps for constructing a table of marketing statistics using a variety of features. Missing values and formats are used to control the table's calculations. Labeling and selected options provide finishing touches for controlling the table's appearance.

INTRODUCTION

Business analysts are often called upon to calculate and present statistics that relate to company performance. From sales and marketing to human resources, managers use data to support virtually every decision. In the 90's and beyond, this is the key to attaining a competitive edge.

This paper will follow a typical credit card marketing campaign from acquisition to a measure of two year profitability. The process will begin with a comparison of PROC MEANS and PROC FREQ to PROC TABULATE for computing response and approval rates.

The next step will compare PROC MEANS to PROC TABULATE in calculating net profit for the first two years.

The following step will compare PROC MEANS and PROC TABULATE in calculating balance attrition for the second year. Balance attrition is defined by the decrease in balance over a fixed period.

Finally, the total performance will be combined to demonstrate the flexibility of PROC TABULATE.

ACQUISITION

The first step to improving the profitability of any acquisition campaign is to perform analysis on the results of a previous campaign. For our analysis, you have a sample of data that represents the initial results and two years of performance from a campaign which originated in the second quarter of 1995. For each record, you know if they responded, were approved, and sum of their marginal finance charges (interest rate minus cost of funds), interchange income and fees for the first two years.

They are coded and labeled as follows:
 Response = 0/1 'Responded'
 Approve = 0/1 'Approved'
 Finance Charges = Amount (in dollars)
 Fees = Amount (in dollars)

In addition, you also have account balances for 6/96 and 6/97 and some select credit bureau information.

The sample data set contains 74,944 observations. This represents all the responses and a 4% sample of the non responders (Weight (SMP_WGT) = 25). The goal of this sampling scheme is to get a relatively equal number of responders and non responders for analysis. The campaign results break down as follows:

	Campaign	Sample	Weight
Non Responders	929,075	37,163	25
Non Approved Responders	10,365	10,365	1
Approved Responders	27,416	27,416	1
Total Mailed	966,847	74,944	

The first goal is to measure the response and approval rates for this campaign. The following code uses PROC FORMAT and PROC FREQ to calculate the rates:

```
PROC FORMAT;
VALUE RESP 1 = 'Responded' 0= 'Did Not Respond';
VALUE APPR 1 = 'Approved' 0 = 'Not Approved'
```

```
PROC FREQ DATA=CAMP0295;
FORMAT RESPONSE RESP. APPROVE APPR.;
WEIGHT SMP_WGT;
TABLE RESPONSE*APPROVE;
TITLE 'Campaign Results - 2nd Quarter 1995';
```

The following output is produced:

Campaign Results - 2nd Quarter 1995

TABLE OF RESPONSE BY APPROVE

RESPONSE	APPROVE		
	Not Appr	Oved	Approved
Frequency			
Percent			
Row Pct			
Col Pct			
-----+-----+-----+-----+			
Did Not Respond	929075	0	929075
	96.09	0.00	96.09
	100.00	0.00	
	98.90	0.00	
-----+-----+-----+-----+			
Responded	10365	27416	37772
	1.07	2.84	3.91
	27.42	72.58	
	1.10	100.00	
-----+-----+-----+-----+			
Total	939431	27416	966847
	97.16	2.84	100.00

In reviewing the table, you can see that it displays all the important information. In the 'Total' column across from 'Percent' you find the response rate is equal to 3.91%. If you want the approval rate as a percent of response, look in the 'Responded' row and the 'Approved' column and read the Row Pct which equals 72.58%. To get the approval rate as a percent of the total mailing, look at the 'Responded' row and the 'Approved' column and read the 'Percent' which equals 2.84%. Assuming all approved accounts become cardholders, you can say that 2.84% of the mailed names became cardholders.

This display may serve a statistician for a quick analysis. However, if you are asked to present these figures to management, these tables can be difficult to interpret. Another method of displaying these values is through the use of PROC MEANS. The following datastep creates new variables which will produce response and approval rates in whole percents. A new variable APPR_RSP is created to measure approval as a percent of response:

```
DATA CAMP0295;
  SET CAMP0295;
  MAIL = 1;
  IF RESPONSE = 1 THEN DO;
    RESP_PCT = 100;
    IF APPROVE = 1 THEN DO;
      APPR_PCT = 100;
      APPR_RSP = 100;
    END;
    ELSE DO;
      APPR_PCT = 0;
      APPR_RSP = 0;
    END;
  END;
  ELSE DO;
    APPR_PCT = 0;
    RESP_PCT = 0;
    APPR_RSP = .;
  END;
RUN;
```

The following PROC MEANS code calculates statistics for response, approval and approval as a % of response. The statistics requested are the sum of the weights (SUMWGT), the sample size (N) and the mean:

```
PROC MEANS DATA = CAMP0295
  MAXDEC=2 SUMWGT N MEAN;
  WEIGHT SMP_WGT;
  VAR RESP_PCT APPR_RSP APPR_PCT;
  TITLE 'Campaign Results - 2nd Quarter 1995';
```

This generates the following output:

Campaign Results – 2nd Quarter 1995

Variable	Sumwgt	N	Mean
RESP_PCT	966847.00	74944	3.91
APPR_RSP	37772.00	37772	72.58
APPR_PXT	966847.00	74944	2.84

This output is much easier to interpret. In addition, the sample size through the use of N, gives the number of records on which the statistics are calculated. This can be useful when determining the validity of the statistics.

The response rates are derived by calculating the weighted mean of the responders (=100) and the nonresponders (=0).

The approval rate as a percent of responders takes advantage of the missing values for all nonresponders. By setting the value of APPR_RSP to missing for nonresponders, those observations are eliminated from the calculation. The result is the weighted mean of the approved (=100) and the non approved (=0) among responders only.

The approval rate as a percent of the amount mailed is derived by calculating the weighted mean of the approved (=100) and the non-approved (=0) where the non-approved also includes the non-responders.

In this example PROC MEANS provides a great deal of information. The main limitation is in controlling the number of decimal places. In order to get the desired precision in the response rates, using MAXDEC=2, you must also have two decimal places in the Sumwgt category. Another limitation is in the inflexibility of the table structure.

To create a table with the desired calculations while controlling the formats and appearance, the optimal choice is to use PROC TABULATE. PROC TABULATE allows you to generate the same statistics as in PROC MEANS while you customize the structure.

To begin the table building process in PROC TABULATE, the following code creates a basic table:

```
PROC TABULATE DATA = CAMP0295;
  VAR RESP_PCT APPR_PCT APPR_RSP;
  WEIGHT SMP_WGT;
  TABLE ( RESP_PCT APPR_PCT APPR_RSP)*MEAN;
  TITLE 'Campaign Results - 2nd Quarter 1995';
```

The tabulate structure is very straight forward. The first line invokes the procedure and names the data set. The VAR statement lists the potential candidates for calculation. The table statement sets up the table structure and defines the calculations.

The PROC TABULATE output shows the means for each group:

Campaign Results – 2nd Quarter 1995

```

| RESP_PCT | APPR_RSP | APPR_PCT |
|-----+-----+-----|
| MEAN     | MEAN     | MEAN     |
|-----+-----+-----|
|      3.91|      72.58|      2.84|
|-----+-----+-----|

```

With the addition of some simple labels and formats, you can provide a greater level of detail. The following highlighted code improves the readability of the table:

```

PROC FORMAT;
  PICTURE PERCNT
    LOW-HIGH '0000009.99%'

```

```

PROC TABULATE DATA = CAMP0295
              FORMAT = PERCNT.;
VAR RESP_PCT APPR_RSP APPR_PCT;
WEIGHT SMP_WGT;
TABLE ( RESP_PCT='Response Rate'
       APPR_RSP='Approval Rate as % of Response'
       APPR_PCT='Approval Rate as % of Total Mailed')
*MEAN=' ';
TITLE 'Campaign Results - 2nd Quarter 1995';

```

The following table is produced:

Campaign Results - 2nd Quarter 1995

```

-----
|           | Approval | Approval |
| Response | Rate as % | Rate as % |
| Rate     | of Response | of Total |
|-----+-----+-----|
|           |           |           |
| 3.91%    | 72.58%    | 2.84%    |
|-----+-----+-----|

```

With the added format option `FORMAT=PERCNT.`, you can increase the width of all cells to 10 spaces while limiting the decimal place to one. Also, note how the `MEAN=' '` eliminates the row containing the word 'MEAN'. The use of the blank label can be used anywhere in the table to eliminate unnecessary labels.

Suppose you also want to display the number mailed and the sample size. You can add these values to the table calculations. You will need to create individual formats for the sums and means. The following code highlights the changes:

```

PROC FORMAT;
  PICTURE PERCNT
    LOW-HIGH '000009.99%'

```

PROC FORMAT is run to reduce the size of the percent format to 9 positions.

```

PROC TABULATE DATA = CAMP0295;
VAR MAIL RESP_PCT APPR_PCT APPR_RSP;
WEIGHT SMP_WGT;
TABLE MAIL='Total Mailed'*SUM=' '*F=COMMA10.
      N='Sample Size'*F=COMMA7.
      ( RESP_PCT='Response Rate'
        APPR_RSP='Approval Rate as % of Response'
        APPR_PCT='Approval Rate as % of Total
        Mailed')
*MEAN=' '*F=PERCNT.;
TITLE 'Campaign Results - 2nd Quarter 1995';
RUN;

```

The following table is produced:

Campaign Results - 2nd Quarter 1995

```

-----
|           |           | Approval |           | |
|           |           | Rate as % | Approval |
| Total    | Sample | Response | of       | Approval |
| Mailed   | Size   | Rate     | Response | Rate as % |
|-----+-----+-----+-----+-----|
|           |           |           |           |           |
| 966,847 | 74,944 | 3.91%    | 72.58%   | 2.84%    |
|-----+-----+-----+-----+-----|

```

As the table becomes more inclusive, you can see how much easier it is to find the desired information. The value in the 'MAIL' column sums the total mailed in the campaign through the use of the weights. This is equivalent to the `Sumwgt` option in PROC MEANS.

ACCOUNT PROFIT - FIRST TWO YEARS

Profits from a credit card account are often calculated over several years to measure income after interest rates and spending patterns are established. This also allows you to determine revolver activity and attrition behavior. Along with other data available at the time of prescreen, this information can be used to improve targeting through better segmentation. The first step is to calculate two year profit using the following code:

```

DATA CAMP0295;
  SET CAMP0295;
  INT_INC = (.18 - .06)/.18*SUM(OF FCG1-FCG24);
  LABEL INT_INC='Interest Income - 2 Years';
  FEE_INC = SUM(OF FEE1-FEE24);
  LABEL FEE_INC='Fee Income - 2 Years';
  IF APPROVE = 1 THEN
    TOT_PROF = INT_INC + FEE_INC - 60;
  ELSE IF RESPONSE = 1 THEN TOT_PROF = -4;
  ELSE TOT_PROF = -1;
  LABEL TOT_PROF='2 Yr Combined Net Profits';
RUN;

```

In the data step, you calculate the incremental interest income by taking the ratio of the cost of funds (.06 = our cost in borrowing) to the total APR (.18 = annual interest rate). By multiplying the total interest income times this ratio, you get the amount of interest margin.

The fee income is simply a total of the fees collected, which represents pure profit to us. The total profit for approved accounts is the sum of the two profit components less the cost of the mailing, application processing and account maintenance for two years (\$60). The total profit for responders is minus the cost of the mailing and application processing (\$4). The total profit for nonresponders is minus the cost of the mailing (\$1).

One method of calculating the average and total profits for any segment is by using PROC MEANS. The following code calculates the total profits of each type and the combined account profit:

```
PROC MEANS DATA = CAMP0295 SUMWGT
      MAXDEC=2 N SUM MEAN;
WEIGHT SMP_WGT;
VAR INT_INC FEE_INC TOT_PROF;
TITLE1 'Campaign Results - 2nd Quarter 1995';
RUN;
```

The output displays the desired information and is fairly easy to interpret.

Two Year Profits – 3rd Quarter 1995 Campaign

Variable	Label	Sumwgt	N
966847.00	74944		
INT_INC	Interest Income – 2 Years	966847.00	74944
TOT_PROF	2 Yr Combined Net Profit	966847.00	74944

Variable	Label	Sum	Mean
FEE_INC	Fee Income – 2 Years	692220.38	0.71
INT_INC	Interest Income – 2 Years	27535763.58	28.39
TOT_PROF	2 Yr Combined Net Profit	25607383.95	26.40

The table displays the sum and average of each type of profit along with the total. For example, the total fee income for the 2nd Quarter 1995 campaign is \$692,220.38 while the average fee income is \$0.71 per account.

There are several limitations to the display through PROC MEANS. Again, the MAXDEC= option is imposed on every value which may not be desirable. For example, in the Sumwgt calculation, MAXDEC=0 would be preferred. In addition, PROC MEANS does not allow for comma and dollar formats.

The following PROC TABULATE code creates a table which includes the sums and means for the profit calculations while allowing for individualized formats:

```
PROC TABULATE DATA = CAMP0295;
WEIGHT SMP_WGT;
VAR MAIL FEE_INC INT_INC TOT_PROF;
TABLE MAIL=Total Mailed*SUM=' *F=COMMA11.
      N='Sample Size'*F=COMMA11.
      (FEE_INC='Total Fee Income'
      INT_INC='Total Interest Income'
      TOT_PROF='Total Combined Net Profit')
      *SUM=' *F=DOLLAR11.
      (FEE_INC='Average Fee Income'
      INT_INC='Average Interest Income'
      TOT_PROF='Average Combined Net Profit')
      *MEAN=' *F=DOLLAR11.2
,ALL=' / RTS=20 ROW=FLOAT;
```

```
TITLE 'Campaign Results - 2nd Quarter 1995';
RUN;
```

The table is structured a little differently to make it more vertical. This is achieved by crossing the desired calculations with the ALL variable. By setting ALL=' ', the heading is eliminated entirely. Some options have been added at the end of the table. RTS=20 forces 20 spaces for the row headings. ROW=FLOAT removes any unnecessary blank spaces in the rows.

The output below shows how the individual labels and formats make it much easier to find the desired information:

Two Year Profits – 3rd Quarter 1995 Campaign

Total Mailed		966,847
Sample Size		74,944
Total Fee Income		\$692,220
Total Interest Income		\$27,535,764
Total Combined Net Profit		\$25,607,384
Average Fee Income		\$0.71
Average Interest Income		\$28.39
Average Combined Net Profit		\$26.40

BALANCE ATTRITION

Another significant driver of profit is balance attrition. In today's balance transfer environment, balance attrition has become much more unstable. Banks are seeking customers who are less likely to transfer balances away from their account. The following data step calculates 2nd year balance attrition for each account:

```
DATA CAMP0295;
  SET CAMP0295;
  BAL_DROP = BAL_0696 - BAL_0697;
  IF BAL_0696 > 0 THEN BAL_ATTR =
    100*BAL_DROP/BAL_0696;
  ELSE BAL_ATTR= 0;
  IF APPROVE = 1 THEN DO;
    BEG_BAL = BAL_0696;
    END_BAL = BAL_0697;
  END;
  ELSE DO;
    BEG_BAL = .;
    END_BAL = .;
  END;
  LABEL BEG_BAL = 'Balance 6/96';
  LABEL END_BAL = 'Balance 6/97';
  LABEL BAL_ATTR = 'Balance Attrition';
```

This data step calculates the amount of balance attrition from the end of the first year to the end of the second year. The first calculation in the data step creates BAL_DROP which represents the change in balance from 6/96 to 6/97. The attrition is calculated as the percent decrease in balance between 6/96 and 6/97. Additional code creates values (including missing values) for BEG_BAL and END_BAL so that the means are calculated only on accounts which were approved.

The following code displays the 6/96 and 6/97 total balances and calculates 2nd year balance attrition using PROC MEANS:

```
PROC MEANS DATA = CAMP0295 SUMWGT
  MAXDEC=2 N SUM MEAN;
  WEIGHT SMP_WGT;
  VAR BEG_BAL END_BAL BAL_ATTR;
  TITLE1 '2nd Year Balance Attrition -
    2nd Quarter 1995 Campaign';
  RUN;
```

The following output is created:

2nd Year Balance Attrition – 2rd Quarter 1995 Campaign

Variable	Label	Sumwgt	N	Sum
BEG_BAL	Balance 6/96	966847.00	74944	60359328.21
END_BAL	Balance 6/97	966847.00	74944	52349027.23
BAL_ATTR	Balance Attrit	27449.00	27449	722590.44

Variable Label Mean

```
-----
BEG_BAL  Balance 6/96      2198.96
END_BAL  Balance 6/97      1907.14
BAL_ATTR Balance Attrition  26.32
-----
```

The values are easy to read. You can see, for example, that in the second year the bank lost \$722,590 in balances from the 2nd Quarter 1995 campaign. Also, the average balance attrition rate is 26.32%.

As in previous PROC MEANS output, the formats are inflexible. In addition, one of the major limitations of PROC MEANS is its inability to calculate means over a group of accounts. In other words, let's say you want to know the 2nd year balance attrition for the entire portfolio. This could easily be calculated by taking the sum of the balances in 6/96, subtracting the sum of the balances in 6/97 and dividing by the balances in 6/96. The equation is: $(60359328 - 52349027)/60359328 = 15.3\%$

This calculation is not possible in PROC MEANS. Instead, PROC MEANS calculates the balance attrition for each individual and averages that value across all accounts. In this example, the average individual attrition is 26.32%. In business, the attrition rate for the group is preferred. PROC TABULATE has the flexibility of calculating attrition either way.

The following code uses PROC TABULATE to calculate the mean balance attrition for each account as well as the overall attrition:

```
PROC FORMAT is rerun to increase the length of the
percent format to 11 positions.

PROC FORMAT;
  PICTURE PERCNT
    LOW-HIGH '0000009.99%'
;
  RUN;
```

```
PROC TABULATE DATA = CAMP0295;
  WEIGHT SMP_WGT;
  VAR MAIL BAL_0696 BAL_0697 BEG_BAL
    BAL_DROP BAL_ATTR;
  TABLE MAIL=Total Mailed*SUM='*F=COMMA11.
    N='Sample Size'*F=COMMA6.
    (BAL_0696='Total Balances 6/96'
    BAL_0697='Total Balances 6/97'
    BAL_DROP='Total 2nd Yr Balance Attrition')*
    SUM='*F=DOLLAR11.
    (BEG_BAL='Average Balance 6/96'
    END_BAL='Average Balance 6/96')*
    MEAN='*F=DOLLAR11.
    BAL_ATTR='Average Balance Attrition (Ind)*
    MEAN='*F=PERCNT.
    BAL_DROP='Average Balance Attrition (Grp)*
    PCTSUM<BAL_0696>='*PERCNT.
```

```
,ALL=' / RTS=20 ROW=FLOAT;
TITLE '2nd Year Balance Attrition -
      2nd Quarter 1995 Campaign';
```

The following table clearly displays the information:

2nd Year Balance Attrition – 2rd Quarter 1995 Campaign

```
-----
-----
|Total Mailed          | 966,847|
|-----+-----|
|Sample Size          | 74,944|
|-----+-----|
|Total Balances 6/96  | $60,359,328|
|-----+-----|
|Total Balances 6/97  | $52,349,027|
|-----+-----|
|Total Dollar Attrition | $8,010,301|
|-----+-----|
|Average Balance 6/96  | $2,199|
|-----+-----|
|Average Balance 6/97  | $1,907|
|-----+-----|
|Percent Balance Attrition (Ind) | 26.3%|
|-----+-----|
|Percent Balance Attrition (Grp) | 15.3%|
|-----+-----|
-----
```

The last two rows display the two calculations for average balance attrition. The first one called 'Average Balance Attrition (Ind)' is the average of the precalculated balance attrition for each account. Of more interest is the value labeled 'Average Balance Attrition (Grp)', which calculates the balance attrition for the portfolio as a whole.

Since targeting decisions are made by selecting profitable groups, the group attrition measure is essential for targeting analysis.

TOTAL PERFORMANCE

Now that we have created tables for all of the values of interest, we will summarize them in one final table with the following code:

```
PROC TABULATE DATA = CAMP0295;
  WEIGHT SMP_WGT;
  VAR MAIL BAL_0696 BAL_0697
      BAL_DROP BAL_ATTR;
```

```
TABLE MAIL=Total Mailed*SUM=' *F=COMMA11.
  N='Sample Size'*F=COMM11.
  (RESP_PCT='Response Rate'
  APPR_RSP='Approval Rate as % of Response'
  APPR_PCT='Approval Rate as % of Total Mailed')
  *MEAN=' *F=PERCNT.
  (FEE_INC='Total Fee Income'
  INT_INC='Total Interest Income'
  TOT_PROF='Total Combined Net Profit')
  *SUM=' *F=DOLLAR11.
  (FEE_INC='Average Fee Income'
  INT_INC='Average Interest Income'
  TOT_PROF='Average Combined Net Profit')
  *MEAN=' *F=DOLLAR11.2
  (BAL_0696='Total Balances 6/96'
  BAL_0697='Total Balances 6/97'
  BAL_ATTR='Total 2nd Yr Balance Attrition ')*
  SUM=' *F=DOLLAR11.
  (BEG_BAL='Average Balance 6/96'
  END_BAL='Average Balance 6/96'
  BAL_ATTR='Average Balance Attrition (Ind)')*
  MEAN='Average'*F=PERCNT.)
  BAL_DROP='Average Balance Attrition (Grp)*'
  PCTSUM<BAL_0696>=' *PERCNT.
  ,ALL=' /RTS=25 ROW=FLOAT;;
TITLE 'Campaign Results - 2nd Quarter 1995';
RUN;
```

Notice the entire table is crossed with 'ALL' to create a vertical display.

The output is displayed on the following page. Again, it is very easy to read. All of the important performance measures are available, from acquisition to profit and balance attrition. With its simplicity and accuracy, this table can serve as a comprehensive report to all levels of management.

Total Performance – 2nd Quarter 1995 Campaign

```
-----
-----
|Total Mailed          | 966,847|
|-----+-----|
|Sample Size          | 74,944|
|-----+-----|
```

Total Responders	37,814
Total Approved	27,449
Response Rate	3.91%
Approval Rate as % of Resp	72.58%
Approval Rate as % of Total	2.84%
Total Fee Income	\$692,220
Total Interest Income	\$27,535,764
Total Combined Profits	\$25,607,384
Average Fee Income	\$0.71
Average Interest Income	\$28.39
Average Combined Profits	\$26.40
Percent Balance Attr (Ind)	26.3%
Percent Balance Attr (Grp)	15.3%

AUTHOR CONTACT

C. Olivia Rud
 DirectCom
 1554 Paoli Pike #286
 Fort Collins, CO 19380

Voice: (610) 918-3801
 Fax: (610) 918-3974
 Internet: OliviaRud@aol.com

SAS is a registered trademark or trademark of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

CONCLUSION

While PROC MEANS and PROC FREQ can handle most of our statistical needs, PROC TABULATE is more powerful. This paper has demonstrated a few of the key situations where PROC TABULATE has superior functionality and flexibility.

REFERENCES

SAS Institute Inc. (1990) *SAS Guide to TABULATE Processing, Second Edition*, Cary NC: SAS Institute Inc.

SAS Institute Inc. (1990) *Procedures Guide, Third Edition*, Version 6, Cary NC: SAS Institute Inc.