USING SAS® TO ASSESS FACTORS AFFECTING EMPLOYEE ATTITUDES AND TURNOVER IN TEMPORARY OFFICES

Mark Holdrege, Bureau of the Census
John Cunningham, Bureau of the Census

ABSTRACT:

The analysis of data collected from separating employees is extremely valuable in understanding the determinants of turnover and in designing work environments. This analysis is particularly valuable when the environments studied are created temporarily every ten years, as is the case with the processing offices (POs) established to process the 1990 Decennial Census of Population and Housing. Personnel Division developed an exit questionnaire for employees separating from these offices to complete and entered the data into Lotus 1-2-3® spreadsheets. We then converted the data to SAS® data sets on the personal computer. We used the tabulate procedure in base SAS® software to produce the initial multiple response reports and used SAS/GRAPH® to produce and present information for preliminary analysis by managers. In the second phase of our analysis, we will extract and match relevant work history data from the National Finance Center which is the host personnel/payroll system for the Bureau of the Census. Using FOCUS™, we will download this data to a print file which we will subsequently convert into a SAS® data set in the VAX™ environment. This second phase will provide the context for a more detailed analysis of responses and will be completed once the temporary PO's close in 1991.

BACKGROUND:

The Bureau of the Census established seven PO’s to expedite the data capture and processing of the 1990 Decennial Census. Six of these offices are located in new, temporary sites and staffed with temporary personnel. The seventh PO is an expansion of our permanent processing facility. This paper describes the application of base SAS® software and SAS/GRAPH® to the analysis of employee information relating to the six temporary offices. The exit questionnaire project is Personnel Division’s first large scale use of SAS® software. We see many future applications for SAS® software and are using this project to gain experience with the system.

Decennial censuses, by definition, occur every ten years. The use of six new processing office facilities constituted a threefold increase over the number of such offices used in processing the 1980 census. Each of the six offices had a peak staffing level of approximately 1800 employees.

We introduced a number of innovations in the pay and staffing systems for the PO employees. To collect and analyze data on employee attitudes toward these systems and their work, we designed an exit questionnaire which all employees complete upon separation from the PO. The employees supply two types of information on the questionnaire. The first type of information consists of identification factors such as PO location and employee's name. The second type of information consists of employee attitudes regarding employment. Identification factors were formatted as character data, and attitudinal responses were formatted as numerical data. This facilitated analysis by base SAS® software and SAS/GRAPH®.

Two analysts independently read all of the responses to the open ended questions for the initial set of approximately 1000 questionnaires received. They identified the categories into which each of the responses could be placed. After comparing results, the analysts agreed to the final set of categories for all of the responses to the questionnaire. This resulted in the creation of 104 variables and formed the basis for the initial coding.

The coding of the numeric variables fell into three categories:

- Single Response Data - We coded the number of the response chosen by the respondent.
- Multiple Response Data - We entered each possible response to the question as a separate variable on the spreadsheet. If the respondent selected that variable in response to the question, we coded the variable as 1, otherwise we left it blank.
- Ordinal Multiple Response Data - We entered each possible response to the question as a separate variable on the spreadsheet. Each variable had a degree of influence attached to it. We coded the degree of influence selected by the respondent (e.g. 1 for "Strong Influence", 2 for "Moderate Influence", 3 for "Not an Influence".)

In order to convert the Lotus 1-2-3® spreadsheets into a SAS® data set, we had to perform two steps. First, we translated each spreadsheet into a Data Interchange Format (DIF) file using the Lotus 1-2-3® translate function. Second, we converted the DIF file into a SAS® data set using the base SAS® DIF procedure.

We encountered a limitation on the size of the spreadsheets that Lotus 1-2-3® could translate to DIF files; therefore, we had to enter the data on more than one spreadsheet. We translated each of these spreadsheets into a DIF file and stored those files on the hard drive of the PC as EXIT1.DIF, EXIT2.DIF, EXIT3.DIF, and so forth.

We processed the conversion of the DIF files into temporary SAS® data sets using the base SAS® DIF procedure as follows.

```
FILENAME NICKNAME = 'EXIT1.DIF';
PROC DIF DIF - NICKNAME OUT = 'FILE1';
RUN;
```

Our initial analysis measures employee attitudes toward pay. The pay rates that we developed were designed to be competitive with comparable rates for administrative and data processing positions across the country. We supplemented this pay structure by establishing a schedule of regular pay increases intended to compensate employees for length of service in the PO. Because employee satisfaction with these pay systems acts as one measure of their effectiveness, we created tables and graphs to display this information.

We are using base SAS® software and SAS/GRAPH®, in part, to evaluate their ability to be used effectively by a staff of Personnel Specialists and Assistants in incorporating and analyzing data from several sources. In addition, we are evaluating our ability to create summaries that provide meaningful information to managers in all areas of administration.

I. Coding Exit Questionnaire Responses

We ask PO employees to complete a four page exit questionnaire upon separation from the PO. The employees supply two types of information on the questionnaire. The first type of information consists of identification factors such as PO location and employee's name. The second type of information consists of employee attitudes regarding employment. Identification factors were formatted as character data, and attitudinal responses were formatted as numerical data. This facilitated analysis by base SAS® software and SAS/GRAPH®.

Two analysts independently read all of the responses to the open ended questions for the initial set of approximately 1000 questionnaires received. They identified the categories into which each of the responses could be placed. After comparing results, the analysts agreed to the final set of categories for all of the responses to the questionnaire. This resulted in the creation of 104 variables and formed the basis for the initial coding.

The coding of the numeric variables fell into three categories:

- Single Response Data - We coded the number of the response chosen by the respondent.
- Multiple Response Data - We entered each possible response to the question as a separate variable on the spreadsheet. If the respondent selected that variable in response to the question, we coded the variable as 1, otherwise we left it blank.
- Ordinal Multiple Response Data - We entered each possible response to the question as a separate variable on the spreadsheet. Each variable had a degree of influence attached to it. We coded the degree of influence selected by the respondent (e.g. 1 for "Strong Influence", 2 for "Moderate Influence", 3 for "Not an Influence").

In order to convert the Lotus 1-2-3® spreadsheets into a SAS® data set, we had to perform two steps. First, we translated each spreadsheet into a Data Interchange Format (DIF) file using the Lotus 1-2-3® translate function. Second, we converted the DIF file into a SAS® data set using the base SAS® DIF procedure.

We encountered a limitation on the size of the spreadsheets that Lotus 1-2-3® could translate to DIF files; therefore, we had to enter the data on more than one spreadsheet. We translated each of these spreadsheets into a DIF file and stored those files on the hard drive of the PC as EXIT1.DIF, EXIT2.DIF, EXIT3.DIF, and so forth.

We processed the conversion of the DIF files into temporary SAS® data sets using the base SAS® DIF procedure as follows.

```
FILENAME NICKNAME = 'EXIT1.DIF';
PROC DIF DIF - NICKNAME OUT = 'FILE1';
RUN;
```
After running the above program for each of the DIF files, we combined the temporary data sets into one permanent SAS® data set. Ordinarily, we could have done this using a SET statement as follows.

LIBNAME PERM 'C:';
DATA PERM.EXIT;
SET FILE1 FILE2 etc...;
RUN;

However, we could not use a simple SET statement as above because of a problem encountered in the conversion process. The method used to code multi-response data and the limit on the number of observations on each spreadsheet resulted in incompatible formats from one data set to the next. This prevented the combination of the temporary data sets into one permanent data set.

The translate function in Lotus 1-2-3® processed a maximum of one hundred observations on any one spreadsheet. By limiting each spreadsheet to a small sample of questionnaires, several columns on the spreadsheet had no responses attributed to them. Hence, the column would appear blank on the spreadsheet.

In the process of converting each spreadsheet into a SAS® data set, the blank numeric columns were assigned a character format rather than the appropriate numeric format. The variables so affected varied from file to file. When we combined the data sets using the SET statement, error messages stated that the same variables appeared in both numeric and character formats. This prevented the SET statement from processing.

We resolved this problem by entering a dummy observation as the first observation on each spreadsheet. We labeled the dummy observation "TEST" in the column reserved for the respondent's last name. We entered a 1 in all columns intended to be numeric. This ensured that all numeric columns would be formatted as such regardless of the presence or absence of actual observations.

By using the following program, we successfully combined the data sets into a permanent data set while also eliminating all of the dummy observations.

LIBNAME PERM 'C:';
DATA PERM.EXIT;
SET FILE1 FILE2 etc...;
IF LNAME = 'TEST' THEN DELETE;
RUN;

Thus, we removed all inconsistencies in the formatting of data and converted the Lotus 1-2-3® spreadsheets into one permanent SAS® data set.

II. Creating Tables To Analyze Pay Data

We selected two variables for our initial measurement of employee attitudes toward pay. The first variable appears in the exit questionnaire as a multiple response variable listing reasons for leaving. Because pay appears among the list of possible factors which contributed to leaving, we can measure the relative importance employees attach to it as a factor in leaving. The second variable appears in the exit questionnaire as a single response variable measuring satisfaction with the schedule of automatic pay increases.

We were interested not only in measuring employee attitudes toward pay but also in observing any relationship between pay and overall job satisfaction. This appears in the exit questionnaire as a single response variable measuring overall job satisfaction. To bring these factors together, we selected the tabulate procedure.

Because we wanted to analyze each pay variable within the context of overall job satisfaction, we proceeded with two separate analyses. The first table cross tabulated reasons for leaving with overall job satisfaction. The second table cross tabulated satisfaction with the system of automatic pay increases with overall job satisfaction.

We established several requirements for our table. First, we wanted to include a non-response tally to identify the data we were failing to capture from our employees. Second, we sought a method of redefining data that we could apply to both single response data as well as multiple response data. We found all of these requirements satisfied in a method described by Tina Keene of the SAS® Institute Inc. in her SUGI 14 paper Analyzing Multiple Response Data Using the TABULATE Procedure.

Our first task involved the coding of the three variables involved in our analysis: reasons for leaving the processing office, satisfaction with the system of pay increases, and overall job satisfaction. The first variable is a multiple response question. We entered each of the twelve possible responses as a separate variable in the data set. For each observation, we coded as 1 each variable selected by the respondent as a reason for leaving. Otherwise, we left the variables blank. Consequently, it was actually composed of a series of variables. The latter two variables are single response questions ranking satisfaction from 1 (highest) to 5 (lowest). We entered each question as a single variable in the data set, coded responses 1 to 5 and labeled the variables PAYSCHED and JOBSAT, respectively.

The data set variables match the exit questionnaire responses as follows:

<table>
<thead>
<tr>
<th>Reasons for Leaving the PO</th>
<th>Questionnaire Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>SALARY</td>
<td>Salary</td>
</tr>
<tr>
<td>PERMANENCE</td>
<td>Seek a more permanent job</td>
</tr>
<tr>
<td>TASKS</td>
<td>Work tasks and responsibilities</td>
</tr>
<tr>
<td>SCHEDULE</td>
<td>Work schedule</td>
</tr>
<tr>
<td>BENEFITS</td>
<td>Health benefits</td>
</tr>
<tr>
<td>SUPERVISOR</td>
<td>Immediate supervisor</td>
</tr>
<tr>
<td>ENVIRONMENT</td>
<td>Work environment</td>
</tr>
<tr>
<td>HEALTH</td>
<td>Health</td>
</tr>
<tr>
<td>MOVING</td>
<td>Transfer out of the area</td>
</tr>
<tr>
<td>PERSONAL</td>
<td>Family or other personal reasons</td>
</tr>
<tr>
<td>SCHOOL</td>
<td>Going back to school</td>
</tr>
<tr>
<td>COMMUTE</td>
<td>Travel to work</td>
</tr>
</tbody>
</table>

Other Variables

<table>
<thead>
<tr>
<th>Data Set Variable</th>
<th>Questionnaire Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAYSCHED</td>
<td>Satisfaction with the Schedule of Pay Increases</td>
</tr>
<tr>
<td>JOBSAT</td>
<td>Satisfaction with Employment</td>
</tr>
</tbody>
</table>

For the multiple response question, we needed to make all of the separate variables facets of one variable while also establishing the non-response tally. For the single response question, we needed to establish the non-response tally. Consequently, we established two new variables. The variable LEAVE incorporates all of the reasons for leaving as well as the non-response tally. The variable INCREASE incorporates the non-response tally with the variable PAYSCHED.

DATA PERM.FILE;
SET PERM.EXIT;
LENGTH LEAVE $16;

The above program establishes a permanent data set containing a new variable, LEAVE. We established a similar but separate program to create a permanent data set containing the new variable, INCREASE.

We used IF-Then statements to combine the reasons for leaving and the non-response tally.

IF SALARY = 1 THEN DO;
LEAVE = 'SALARY';
OUTPUT;
END;
IF PERMANENCE = 1 THEN DO;
LEAVE = 'PERMANENCE';
OUTPUT;
END;
•
•
IF COMMUTE = 1 THEN DO;
LEAVE = 'COMMUTE';
OUTPUT;
END;
IF SALARY =., AND PERMANENCE =. AND...COMMUTE =.
THEN DO;
LEAVE = '* NO RESPONSE *';
OUTPUT;
END;
In a similar manner, we used If-Then statements to combine
PAYSCHED with the non-response tally.
IF PAYSCHED = 1 THEN DO;
INCREASE = 'VERY GOOD';
OUTPUT;
END;
•
•
IF PAYSCHED = 5 THEN DO;
INCREASE = 'VERY POOR';
OUTPUT;
END;
IF PAYSCHED = . THEN DO;
INCREASE = '* NO RESPONSE *';
OUTPUT;
END;

FIGURE 1.
REASONS FOR LEAVING THE PROCESSING OFFICE
WITHIN THE CONTEXT OF OVERALL JOB SATISFACTION

<table>
<thead>
<tr>
<th>REASONS FOR LEAVING THE PROCESSING OFFICE</th>
<th>SATISFACTION WITH AUTOMATIC PAY INCREASES</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WITHIN THE CONTEXT OF OVERALL JOB SATISFACTION</td>
</tr>
<tr>
<td></td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>GOOD</td>
</tr>
<tr>
<td>* NO RESPONSE *</td>
<td>13</td>
</tr>
<tr>
<td>BENEFITS</td>
<td>23</td>
</tr>
<tr>
<td>COMMUTE</td>
<td>9</td>
</tr>
<tr>
<td>DUTIES</td>
<td>1</td>
</tr>
<tr>
<td>ENVIRONMENT</td>
<td>1</td>
</tr>
<tr>
<td>HEALTH</td>
<td>13</td>
</tr>
<tr>
<td>PERMANENCE</td>
<td>66</td>
</tr>
<tr>
<td>RELOCATION</td>
<td>19</td>
</tr>
<tr>
<td>SALARY</td>
<td>21</td>
</tr>
<tr>
<td>SCHEDULE</td>
<td>27</td>
</tr>
<tr>
<td>SCHOOL</td>
<td>19</td>
</tr>
<tr>
<td>SUPERVISOR</td>
<td>0</td>
</tr>
<tr>
<td>TOTALS</td>
<td>257</td>
</tr>
</tbody>
</table>

Our final step in redefining the data entailed using a keep
statement to retain the class variables to be used in our table:

KEEP LEAVE JOBSAT;

The following program creates the table presented in Figure
1. The table consists of a frequency distribution between the
reasons for leaving and overall job satisfaction.

OPTION PS=60 NODATE NONNUMBER;
PROC FORMAT;
VALUE JOBSATFMT 1 = 'VERY GOOD' 2 = 'GOOD'
3 = 'FAIR' 4 = 'POOR' 5 = 'VERY POOR';
PROC TABULATE DATA=PERM.FILE FORMAT = 4.0;
CLASS LEAVE JOBSAT;
TABLE LEAVE ALL,JOBSAT ALL/NRTS=17 ROW=FLOAT
MISSSTEXT='O' BOX='REASONS FOR LEAVING THE PROCESSING OFFICE';
FORMAT JOBSAT JOBSATFMT;
LABEL LEAVE='REASONS' JOBSAT='SATISFACTION WITH
EMPLOYMENT';
KEYLABEL ALL='TOTALS';
TITLE1 REASONS FOR LEAVING THE PROCESSING OFFICE';
TITLE2 'WITHIN THE CONTEXT OF OVERALL JOB SATIS-
FACTION';

We used the format procedure to appropriately label the job
satisfaction coding. We did not need it for the LEAVE variable
because we, in essence, established the labels in the If-Then
statements.

Initial review of the table reveals relatively low item nonre-
response and an apparent divergence in priorities between the
respondents who rated overall job satisfaction as poor or very
poor and respondents who rated overall job satisfaction as fair
to very good.

The table presenting satisfaction with the system of auto-
matic pay increases within the context of overall job satisfaction
appears in Figure 2. Of immediate concern in this table was the
relatively high level of item nonresponse. Based on this initial
information, we modified our instructions to the PO's and held
follow-up discussions with PO managers in an attempt to ensure
the completion of exit questionnaires.

FIGURE 2.
SATISFACTION WITH AUTOMATIC PAY INCREASES
WITHIN THE CONTEXT OF OVERALL JOB SATISFACTION

<table>
<thead>
<tr>
<th>SATISFACTION WITH PAY INCREASES</th>
<th>SATISFACTION WITH AUTOMATIC PAY INCREASES</th>
</tr>
</thead>
<tbody>
<tr>
<td>WITHIN THE CONTEXT OF OVERALL JOB SATISFACTION</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>GOOD</td>
<td>GOOD</td>
</tr>
<tr>
<td>PAY INCREASES</td>
<td>41</td>
</tr>
<tr>
<td>VERY GOOD</td>
<td>81</td>
</tr>
<tr>
<td>GOOD</td>
<td>25</td>
</tr>
<tr>
<td>FAIR</td>
<td>36</td>
</tr>
<tr>
<td>POOR</td>
<td>4</td>
</tr>
<tr>
<td>VERY POOR</td>
<td>3</td>
</tr>
<tr>
<td>TOTALS</td>
<td>190</td>
</tr>
</tbody>
</table>

1058
III. Using SAS/GRAPH® to Analyze Pay Data

The graphics capability of SAS® makes it well suited to prepare summary information for managers. We chose to create stacked bar graphs and pie charts to illustrate and summarize the initial pay data.

We decided to keep the analysis within the framework of overall job satisfaction. To point up the similarities and differences observed between the satisfied and dissatisfied employees, we reduced the job satisfaction categories from five to two. We grouped in the "Satisfied" category all respondents who rated job satisfaction fair to very good. We grouped in the "Dissatisfied" category all respondents who rated job satisfaction poor or very poor.

We regrouped the data by creating a new variable called CLASS and saved the data needed to create the graphs in the temporary data set named REGROUP.

```sas
DATA REGROUP;
SET PERM.FILE;
LENGTH CLASS $30;
IF JOBSAT=1 OR JOBSAT=2 OR JOBSAT=3 THEN DO;
   CLASS = 'SATISFIED';
   OUTPUT;
END;
IF JOBSAT=4 OR JOBSAT=5 THEN DO;
   CLASS = 'DISSATISFIED';
   OUTPUT;
END;
KEEP LEAVE CLASS;
```

To illustrate the reasons for leaving, we created two pie charts, one representing the satisfied employees and the other representing dissatisfied employees.

```sas
DATA SATIS;
SET REGROUP;
WHERE CLASS = 'SATISFIED';
FILENAME TEMP 'A:LEAVE.GSF';
GOPTIONS DEV = HP7475A GSFMODE = REPLACE GSFNAME = TEMP;
PROC GCHART DATA=SATIS;
PIE LEAVE discrete type = PERCENT other = 7 slice = ARROW;
TITLE1 'Reasons for leaving the Processing Office';
TITLE2 'Satisfied Employees';
RUN;
```

The above program creates a pie chart based on the LEAVE variable. We created a temporary data set of satisfied employees which we labeled SATIS. The pie chart for satisfied employees is based on this data set.

To keep the number of slices to a minimum, we combined all responses of less than seven percent into the category 'other'. We present the resulting chart for satisfied employees in Figure 3 and the comparable chart for dissatisfied employees in Figure 4.

A comparison of the two charts points up the differences we noted in the table in Figure 1. The priorities appear to be quite different for the two groups; however, both groups of employees appear to rate pay at approximately the same level.

We produced a stacked bar graph to display levels of employee satisfaction with the system of pay increases.

```sas
DATA REGROUP1;
SET REGROUP;
LENGTH CLASS $30;
IF JOBSAT=1 OR JOBSAT=2 OR JOBSAT=3 THEN DO;
   CLASS = 'SATISFIED';
   OUTPUT;
END;
IF JOBSAT=4 OR JOBSAT=5 THEN DO;
   CLASS = 'DISSATISFIED';
   OUTPUT;
END;
KEEP LEAVE CLASS;
```

IV. Moving Data From the PC to VAX™

Since PO operations are coming to a close and the temporary employees are starting to be released, we are experiencing a substantial increase in the number of exit questionnaires received. This has also created a need for increased storage capacity for the data set. We have, therefore, transferred the exit questionnaire data set from PC to VAX™.
We accomplished the transfer from PC to VAX™ using the upload procedure. This first required establishing the micro-to-host link using a script developed by the SAS® Institute for the PC to VAX™ link.

Once we established this link, we had to define the libraries to be used in each system. For the PC, we submitted the following libname statement, identifying the hard drive:

```
LIBNAME PERM 'C:';
```

For the VAX™, we used RSUBMIT at the command line to define the VAX™ directory in which we wanted to store the SAS® data set:

```
LIBNAME VAX '[VAXDIR]';
```

Having identified the relevant libraries, we used the upload procedure to copy the exit questionnaire data set from the PC to the VAX™.

```
PROC UPLOAD
DATA = PERM.EXIT
OUT = VAX.EXIT;
RUN;
```

Again, we used RSUBMIT at the command line. RSUBMIT is necessary to both assign the libname to the VAX™ directory and upload the data set because these procedures execute in the VAX™ system rather than on the PC.

V. Phase II

We are continuing to collect exit questionnaire data from the processing offices. By the end of 1991, all of the temporary offices will have completed operations, and all temporary employees will have separated. We will then capture and merge downloaded data from our automated personnel/payroll database which the National Finance Center (NFC) of the Department of Agriculture is currently building. We will use the data from the NFC system to provide work and pay history information which will make a more detailed analysis of the exit questionnaire data possible.

This will permit us to place our analysis of employee attitudes within the context of the employees' work histories in the PO's. For instance, based on their personnel records, we will be able to group the respondents into three categories: voluntary separations (resignations), terminations for performance/conduct, and separations due to completion of assignment. These groupings will allow us to test for differences in measured attitudes among and within the various separation categories. We will also be able to place responses within the context of the respondents' educational background, length of service at separation, census work and pay history (e.g., positions held, salary increases, and so forth), and other factors of interest.

We will download the NFC employment data from the IBM® mainframe in New Orleans by use of FOCUS™ language procedures. We will then convert the data to a SAS® data set and match to each of the respondents in the exit questionnaire data.
set. We will use the new data set to summarize and evaluate the attitudinal data for presentation to management and for use in the planning of future census operations.

VI. Summary

In order to better plan future census operations, we developed an exit questionnaire for separating employees. The information on the exit questionnaire was captured by keying it into Lotus 1-2-3 spreadsheets. The spreadsheets were translated to DIF files and then converted to SAS® data sets. We used base SAS® software on PC's for data manipulation and analysis and SAS/GRAPH® for preliminary data presentation. The data sets have been moved to a VAX® minicomputer to permit more extensive analysis and matching to a personnel/pay history data set. The SAS® system software has proved to be an effective and flexible tool for processing and analyzing personnel data.

Acknowledgements

The authors wish to thank Robert Bateman, Michael Bretz, and William Page of the Bureau of the Census for their technical review and assistance.

References

Analyzing Multiple Response Data Using the TABULATE Procedure by Tina Keene of the SAS® Institute, Inc.

SAS and SAS/GRAPH are registered trademarks of the SAS Institute Inc., Cary, NC.

Lotus 1-2-3 is a registered trademark of the Lotus Development Corporation, Cambridge, MA.

IBM is a registered trademark of the International Business Machines Corporation, Armonk, NY.

VAX is a trademark of Digital Equipment Corporation, Maynard, MA.

FOCUS is a trademark of Information Builders, Inc., New York, NY.