SAS project

EXAM OF STATISTICAL SOFTWARE

UNIVERSITY OF BOLOGNA

SUSANNA CASARINI

DATA:

ADDRESSES dataset:
- 424 observations, 9 variables.
- Informations about employee addresses.

ORGANIZATIONS dataset:
- 424 observations, 4 variables.
- Informations about job role and department in the organization.

DONATIONS dataset:
- 124 observations, 7 variables.
- Informations about employee donations, divided in four periods.

AIMs:

- Create a rtf report with employees informations.
- Create a second report with employee informations about donations, dividing employees by department and job role.
- Create a table and a report with informations about donations in a specific period, dividing employees by department.
- Create a second table with the same informations, dividing employees by job role.
- Create a new variable to classify employees as "benefactors" if the total number of donations is greater than 1 and the total amount exceeds 30$.
- Study the association between job role and being a benefactor.

First report: Employee informations.
Records of clients that made at least one donation.

- DATA STEP to create three datasets in the work library
- PROC CONTENTS to understand the datasets
- PROC SORT to sort the data by Employee_ID
- MERGE statement in the data step, to create a new dataset using a one by many procedure, selecting only the ones who made donations.
- CATX statement to create the new variable addresses
- PROC REPORT step to create a rtf file

First report:
Employee informations.
LIBNAME
and PROC CONTENTS

```sas
/*Import the single datasets on SAS*/
libname data "C:\Users\susanna.casarini\Downloads\Data\Data";

/*DATA step to create three datasets*/
data addresses;
      set data.employee_addresses;
run;

data donations;
      set data.employee_donations;
run;

data organization;
      set data.employee_organization;
run;

/*PROC CONTENTS step to study the datasets*/
proc contents data=addresses;
run;

proc contents data=donations;
run;

proc contents data=organization;
run;
```

First report: Employee informations.
PROC SORT, MERGE and CATX statements

```sas
proc sort data=addresses;
      by Employee_ID;
 run;

proc sort data=donations;
      by Employee_ID;

 run;proc sort data=organization;
      by Employee_ID;
 run;
```

```sas
/*DATA step with MERGE statements*/
data employee1;
      merge addresses(in=add)
               donations(in=don)
               organization(in=org);
      by Employee_ID;
      if add=1 AND (qtr1>0 or qtr2>0 or qtr3>0);
 run;

/*CATX statement to create the address variable*/
data employee2;
      set work.employee1;
      address = catx(", ",street_number,street_name,city,country);
 run;

/*PROC REPORT using ODS language to create a first report*/
ods rtf file='C:\Users\susanna.casarini\Downloads\Data\Data\Employees_First_Report.rtf';
proc report data=employee2;
      columns Employee_ID Department Address;
      define employee_ID / display 'Employee ID';
      define department / display 'Department' ;
      define address / display 'Address' ;
      endcomp;
 run;
ods rtf close;
```

First report: Employee informations.
rtf report

Employee ID	Department	Address
120267	Secretary of the Board	43, Grist Mill Rd, San Diego, US
120269	Strategy	65, Applecross Cir, Philadelphia, US
120270	Concession Management	1916, Hebron Church Rd, Miami-Dade, US
120271	Concession Management	94, Irelan Dr, Miami-Dade, US
120272	Concession Management	461, Kimberly Dr, Philadelphia, US
120275	Concession Management	56, Cahill Rd, San Diego, US
120660	Logistics Management	4156, Hwy 42, Miami-Dade, US
120662	Logistics Management	2565, Bledsoe Ave, Miami-Dade, US
120663	Logistics Management	4407, Fountain Dr, Miami-Dade, US
120668	Logistics Management	744, Farmington Grove Dr, Miami-Dade, US

Second report: Employee donations.
Employee donations by department and job role.

- DATA STEP to calculate the mean, the minimum and the maximum of the donations in different periods
- PROC REPORT step to create a rtf file with the statistics of donations for every period, divided by department and job role.

Second report: Employee donations.
DATA step and PROC REPORT

```
/*DATA step to create the three variable: mean min max for the donations in the four periods*/
data employee3;
      set work.employee2;
      don_mean = mean(qtr1, qtr2, qtr3, qtr4);
      don_max = max(qtr1, qtr2, qtr3, qtr4);
      don_min = min(qtr1, qtr2, qtr3, qtr4);
run;

ods rtf file='C:\Users\susanna.casarini\Downloads\Data\Data\Employees_Second_Report.rtf';
proc report data=employee3;
            columns Employee_ID Department Job_title don_mean don_max don_min;
            define employee_ID / display 'Employee ID';
            define department / order 'Department' ;
            define Job_title / order 'Job Title' ;
            define don_mean / display 'Don_Mean' ;
            define don_max / display 'Don_Max' ;
            define don_min / display 'Don_Min' ;
      endcomp;
run;
ods rtf close;
```

Second report: Employee donations.
rtf report of employee donations by department and job role.

Employee ID	Department	Job Title	Don_Mean	Don_Max	Don_Min
120746	Accounts	Account Manager	20	20	20
120752		Accountant I	10	10	10
120761			10	10	10
120762			5	5	5
120754		Accountant II	10	10	10
120759			13.333333	20	5
120757		Accountant III	10	15	5
120748		Building Admin. Manager	20	20	20
120747		Financial Controller I	20	20	20
120753		Financial Controller II	25	25	25
120760		Financial Controller III	17.5	20	15
120767	Accounts Management	Accountant I	15	15	15
120768		Accountant II	20	20	20
120771			20	20	20
120766		Auditing Manager	25	25	25

First table: means of employee donations.
means of donations by department for every period.

- PROC MEANS and CLASS statement to calculate the donations mean for every period by department
- PROC SORT to sort data by department
- PROC TRANSPOSE to transpose the donations variables and create the period variable
- PROC REPORT to display the table

First table: means of employee donations.
PROC MEANS and PROC TRANSPOSE

```
ods rtf file='C:\Users\susanna.casarini\Downloads\Data\Data\Employees_Table.rtf';
proc means data=employee3 min max mean noobs;
      var qtr1 qtr2 qtr3 qtr4 ;
      class department ;
      output out= employee_dep;
run;
ods rtf close;

proc sort data=employee_dep out=employee_sort;
by department;
run;

proc transpose data=employee_sort out=employee_transp name=period;
var qtr1 qtr2 qtr3 qtr4 ;
by department;
run;

proc sort data=employee_transp out=transp2;
by department period;
run;

proc contents data=transp2;
run;
```

First table: means of employee donations.
Table report

Department	Period	Minimum	Maximum	Mean
Accounts	Qtr1	0	25	11.818182
	Qtr2	0	25	12.727273
	Qtr3	0	25	13.636364
	Qtr4	0	25	10.454545
Accounts Management	Qtr1	15	25	19
	Qtr2	15	25	19
	Qtr3	15	25	19
	Qtr4	0	25	15
Administration	Qtr1	0	20	10.5
	Qtr2	0	20	10.5
	Qtr3	5	20	11.5
	Qtr4	5	20	11.5
Concession Management	Qtr1	10	20	16.25
	Qtr2	10	20	13.75

Second table: means of employee donations.
means of donations by job role for every period

- MACRO to use the code of the previous table, setting the grouping variable as option.
- Apply the MACRO to the job role variable

Second table: means of employee donations.
means of donations by job role for every period

```
%macro report(class_used=) ;

    ods rtf file='C:\Users\susanna.casarini\Downloads\Data\Data\Employees_Fourth_Report.rtf';
    proc means data=employee_job min max mean noobs;
        var qtr1 qtr2 qtr3 qtr4 ;
        class &class_used ;
        output out= employee_job;
    run;
    ods rtf close;
```

```
        endcomp;
    run;
    ods rtf close;

%mend report;

%report(class_used=department);

/*Repeat the same operation grouping by job role (use a macro to improve efficiency)*/
%report(class_used=job_title);
```

Second table: means of employee donations.
means of donations by job role for every period

Department	Period	Minimum	Maximum	Mean
Accounts	Qtr1	10	25	16.25
	Qtr2	10	25	17.5
	Qtr3	5	25	15
	Qtr4	5	25	14.375
Accounts Management	Qtr1	15	25	19
	Qtr2	15	25	19
	Qtr3	15	25	19
	Qtr4	15	25	18.75
Administration	Qtr1	5	20	11.666667
	Qtr2	5	20	11.666667
	Qtr3	5	20	11.5
	Qtr4	5	20	11.5
Concession Management	Qtr1	10	20	16.25
	Qtr2	10	20	13.75

Frequency table: benefactor and job role.
Association between being a benefactor and job.

- DATA STEP to create a flag variable BNFL. Classification of "Benefactors" as employees who made at least one donation and total amount exceeds 30$.
- PROC FREQ to study the association between the two variables

Frequency table: benefactor and job role.
PROC FREQ

```
/*Create the variable COUNT_DON that counts the number of donations made.
If the total number of donations is greater than 1 and
the total amount exceed 30$, then categorize the subject as benefactor
(create a flag variable BNFL that assumes value "Y" if the subject is a benefactor and "N"

data employee_count;
    set employee3;
    count_don = 4-(sum(qtr1=0, qtr2=0, qtr3=0, qtr4=0));
    if count_don > 0 AND (qtr1+qtr2+qtr3+qtr4) > 30  then BNFL="Y";
    else BNFL="N";
run;

/*Which role is associated to being benefactor?*/
proc freq data=employee_count;
    table job_title*BNFL / nopercent norow nocol;
run;

proc freq data=employee_count;
    table department*BNFL / nopercent norow nocol;
run;
```

Frequency table: benefactor and job role.
Association between being a benefactor and job.

La procedura FREQ

Tabella di Job_Title rispetto a BNFL

Job_Title	BNFL N	BNFL Y	Totale
Account Manager	0	1	1
Accountant I	2	2	4
Accountant II	2	2	4
Accountant III	1	0	1
Administration Manager	0	1	1
Applications Developer I	0	1	1
Applications Developer IV	2	0	2
Auditing Manager	0	1	1
Auditor I	0	1	1
BI Specialist II	1	0	1
Building Admin. Manager	0	1	1

Department	BNFL N	BNFL Y	Totale
Accounts	5	6	11
Accounts Management	1	4	5
Administration	3	7	10
Concession Management	1	3	4
Engineering	1	1	2
Group HR Management	1	4	5
IS	4	7	11
Logistics Management	2	3	5
Marketing	1	4	5
Purchasing	3	2	5
Sales	17	27	44
Sales Management	0	4	4
Secretary of the Board	0	1	1
Stock & Shipping	3	6	9
Strategy	0	1	1
Totale	42	80	122