



ACC DAR Adapting to Climate Change in Coastal Dar es Salaam

Technical Workshop on Data Analysis

DATA ANALYSIS

Dar es Salaam, 30 September 2011



CENTRO INTERUNIVERSITARIO
DI RICERCA PER LO SVILUPPO
SOSTENIBILE - CIRPS



SAPIENZA
UNIVERSITÀ DI ROMA



Pietro DEMURTAS

IRPPS – National Research Council

p.demurtas@irpps.cnr.it

Program:

- How the sample was chosen? Write a little tale of your experience.
- Review of data treatment procedures.
- Some procedure of data analysis (cross tabs, compare means and multiple responses treatment)



Cross Variables

- The relation between two categorical variables* can be described using contingency tables.

Religion	Vote				
	democratic	republican	Independent	No responses	Total
Catholic	202	248	134	96	680
Protestant	158	12	7	38	215
Jewish	26	2	4	4	36
others	9	1	1	13	45
nothing	15	12	6	12	45
total	410	275	152	163	1000

***REFRESH!!!**

If you have a cardinal variable you can transform it in a categorical variable with SPSS command "RECODE" (i.e. transform age in classes of age)



How do you read a cross tab?

Through the comparison between cells in the contingency table and the marginal (totals) of rows or columns.



A table showing row percentages is a table where marginal rows have a value of 100 and marginal columns are the percentages of column categories on the total.

Religion	Vote				
	democratic	republican	independent	No responses	Total
Catholics	29,7	36,5	19,7	14,1	100
Protestants	73,5	5,6	3,3	17,7	100
Jewish	72,2	5,6	11,1	11,1	100
other	20,0	2,2	2,2	28,9	100
nothing	33,3	26,7	13,3	26,7	100
total	41,0	27,5	15,2	16,3	100



A table showing column percentages is a table where marginal columns have a value of 100 and marginal rows are the percentages of rows categories on the total.

Vote	Religion				
	Catholic	Protestant	Jewish	others	total
Democratic	49,3	90,2	88,2	58,9	68,0
Republican	38,5	4,4	4,6	23,3	21,5
independent	6,3	0,7	2,6	2,5	3,6
no responses	2,2	0,4	0,7	8,0	4,5
nothing	3,7	4,4	3,9	7,4	4,5
Total	100	100	100	100	100



When you would observe the influence that the variable on column has on the variable on row, you must ask to SPSS a cross table with column percentages (and *vice versa*).

In this case:

In the column we have the **Independent Variable**

In the row we have the **Dependent Variable**

*** In our example, the religion of a person (Ind. V.) influences his decision of vote (Dep. V.)**



CROSSTABS

```
/TABLES=vote BY religion  
/FORMAT=AVALUE TABLES  
/CELLS=COUNT COLUMN  
/COUNT ROUND CELL.
```

*in this program we ask to SPSS a cross table with variable 'vote' on row and variable 'religion' on column.

the sub-command FORMAT asks to SPSS that tables are sort in increasing (default command)

the sub-command CELLS asks to insert on table cells combined frequencies (between variables 'vote' and 'religion') and column percentages

the last sub-command asks to insert on table cells absolute values

** we can insert also a sub-command for charts or we can design a chart directly on excel.



Compare Means (only for Cardinal variables)

The procedure "MEANS" allows you to compare the average values of one or more cardinal variables within categories of one or more categorical variables

For example, we use the procedure "MEANS" if we want to compare the wage average of males and females!!!



On SPSS:

MEANS

```
TABLES= wage BY sex BY educational level)  
/CELLS MEAN COUNT STDDEV.
```

- * The keyword MEANS indicates the procedure for comparing the averages.
- ** After the sub-command TABLES , we list the cardinal variables to be compared within the categories of categorical variables. We write the cardinal variables before "BY" and categorical variable after "BY".
- *** The sub-command /CELLS asks the average and number of cases for each group



Multiple responses

In your questionnaire there are questions that can involve more responses.

For example: **Q16. WHAT CROP DO YOU GROW?**

One person will give more answers to this question. For this reason on SPSS there will be more variables for question 16: there will be one variable for each crop grown and each variable will have two values:

0 = not selected and 1=selected



How to analyze multiple responses?

FREQUENCIES:

```
MULT RESPONSE GROUPS=$multi_16 'crop grown' (q16_1 q16_2 q16_3 q16_4 q16_5  
q16_6 q16_7 q16_8 q16_9 q16_10 q16_11 q16_12 q16_13 q16_14 (1))  
/FREQUENCIES=$multi_16.
```

- MULT RESPONCE is the procedure that we want
- After the sub-command GROUPS= there is the name and the label of our synthetic variable.
- Within the brackets there is a list of variables that will be synthesized.
- At the end of this list, into other brackets, there is the value that will be counted (1)



Crosstabs

```
MULT RESPONSE GROUPS=$multi_16 'crop grown' (q16_1 q16_2 q16_3 q16_4 q16_5  
q16_6 q16_7 q16_8 q16_9 q16_10 q16_11 q16_12 q16_13 q16_14 (1))
```

```
/VARIABLES= age(1 4)
```

```
/TABLES=age BY $multi_16
```

```
/CELLS=COLUMN
```

```
/BASE=CASES.
```

- For the first command, see the slide before
- After the sub-command /VARIABLES=, we specify the variable AGE (that will cross with “crop grown”) . Into the brackets we must specify the first and the late value of this variable (in this case is 1 and 4, because the variable age have 4 values).
- After the sub-command /TABLES= we ask to SPSS the contingency table between Age and \$multi_16
- The sub-command /CELLS= asks percentages of column
- On the last row we ask to calculate percentages basing on the Cases. Alternatively we can ask to calculate percentages on Responses. (in the first case probably I will have a percentage like 200%, 300%... Because each person give more than one answer. In the second case the sum of percentages will be as always 100%, because SPSS will compute the percentage on the total of answers)



How to analyze multiple responses?

From the menu:

Analyze



Multiple Response



Define variable Sets:

- To Transfer variables that will be synthesized into the window 'variables in set'
- Define the value that will be counted (for dichotomies) or the range of values (for variable that have more of two values)
- Choice the name and the label of synthetic variable
- Click on "ADD"
- Close the window (...)



(...) Frequencies

Repeat the procedure: From ANALYSIS to MULTIPLE RESPONSE , then you click on Frequencies. You will find the synthetic variable that you have created before. You must transfer this variable into the window named “Table(s) for:” and then you click “OK”.

Cross tabs

Repeat the procedure: from ANALYSIS to MULTIPLE RESPONCE, then you click on Cross tabs. Here you can cross your synthetic variable with others variables and choice percentages of row or column.

When you choice one of others variables, you have to specify the first and the late values of this variable

(if the variable is have 14 values, you must write 1 and 14).

