vam
food security analysis

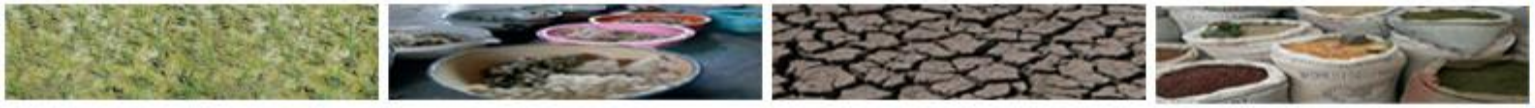
Practicing with SPSS

Bangkok, 6-10 May 2013



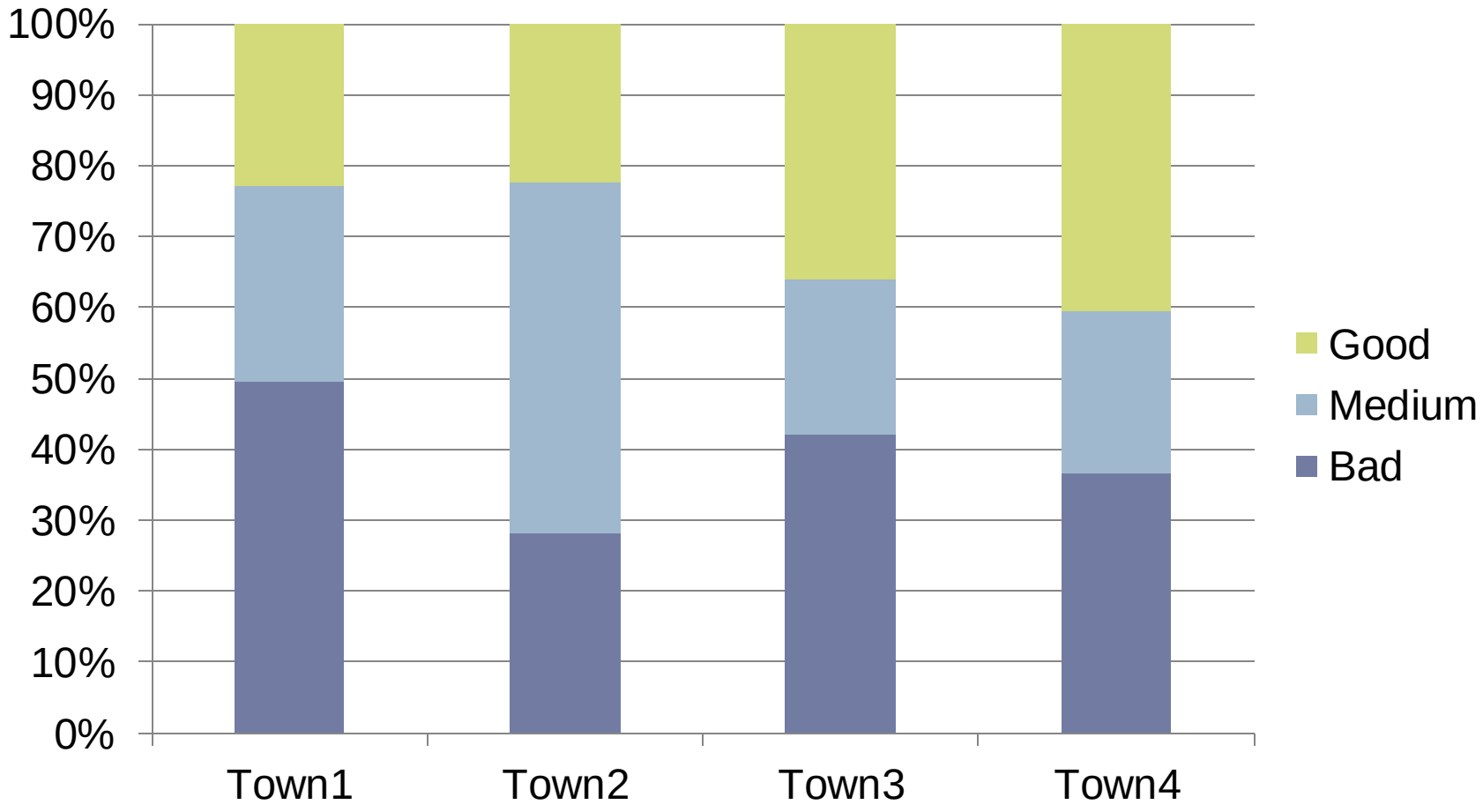

vam
food security analysis

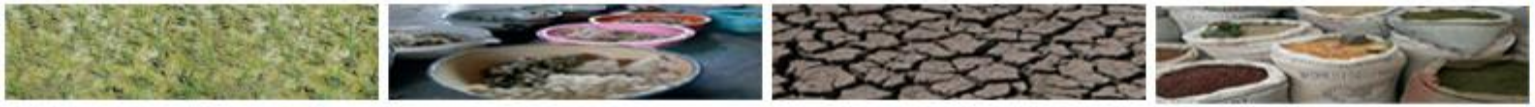
Quiz



vam
food security analysis

How many variables do we need to create this graph?

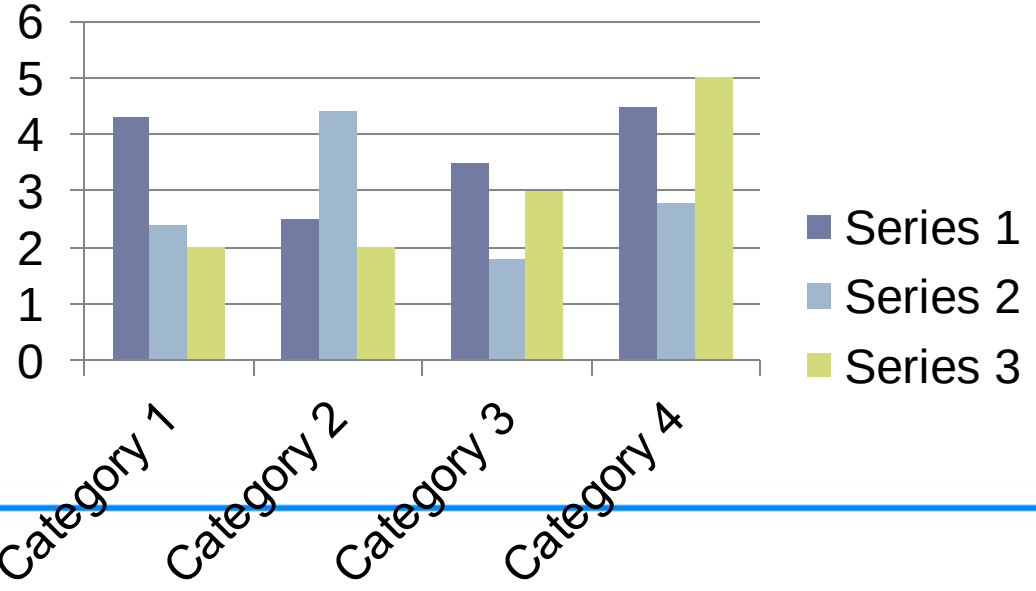




vam food security analysis

Which are the differences between the graph and the table below?

	Series 1	Series 2	Series 3
Category 1	4.3	2.4	2
Category 2	2.5	4.4	2
Category 3	3.5	1.8	3
Category 4	4.5	2.8	5






vam food security analysis

How many variables will come out from this questionnaire?

6.6	Please indicate if you have received **remittances** from the following sources **over the past two years (Oct 09 – Oct 07)** ***Circle all that apply***		1 = Migrant household member (seasonal <1 year) 2 = Migrant household member (long-term >1year)	3 = Other source (non-HH-member) **9 = None of the above** → skip to section VII
6.7	If you have received remittances, please indicate if the amount has **changed** or remained the same **over those two years 2008 and 2009.**		1 = Increased 2 = Decreased 3 = Remained about the same → skip to section VII	
6.8	What is the main reason for this change?	If increased 1 = More members migrating 2 = More employment opportunities 3 = Higher wages/salaries 4 = Fewer members migrating	If decreased 5 = Fewer employment opportunities 6 = Loss of jobs 7 = Lower wages/salaries ***9 = Other***	

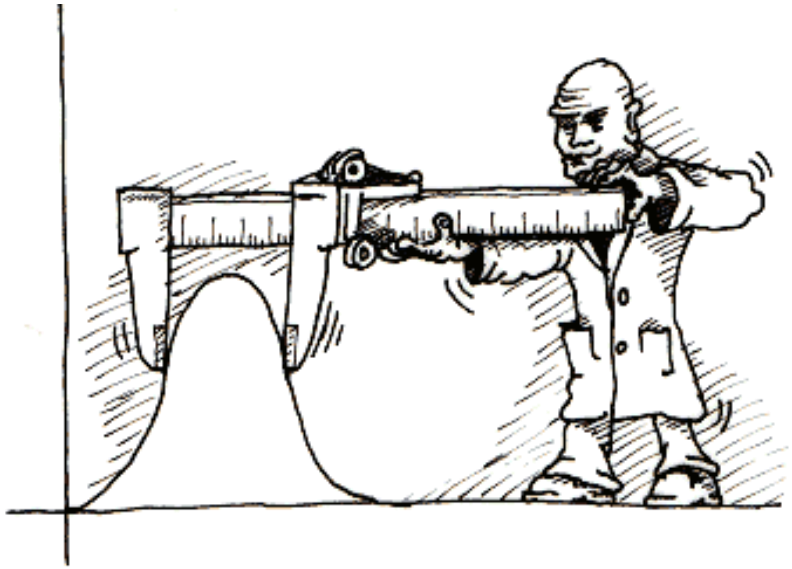
6.3 At what time of the year, did HH members **mainly** work outside of the community **over past 12 months**?

Indicate timing of migration for the HH member whose contribution to the overall HH's income is most important. **Tick all that apply.**

Oct 08	Nov 08	Dec 08	Jan 09	Feb 09	Mar 09	Apr 09	May 09	Jun 09	Jul 09	Aug 09	Sep 09	Oct 09	Nov 09	All

Which is a variable and which is a case?

- Household
- Number of days meat was consumed
- Child
- Mother
- Household size
- Sex of the household head



TYPES OF DATA AND DESCRIPTIVE STATISTICS

Introduction

In the social sciences we are usually interested in **discovering something** about a phenomenon.

Whatever the phenomenon we desire to explain, we seek to explain it by **collecting data** from the real world and then **using these data to draw conclusions** about what is being studied.

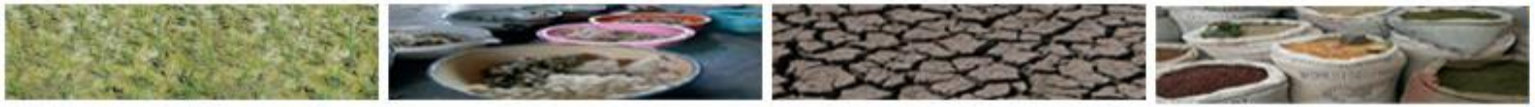
Variable vs. Cases

- **Case**

  statistical unit that is being described (household, people, sampled universe)

- **Variable**

  an indicator that describes a case (sex of the household head, age of the person, income activity of the household)



vam food security analysis

Type of variables

We work with two types of variables

Continuous (Scale)

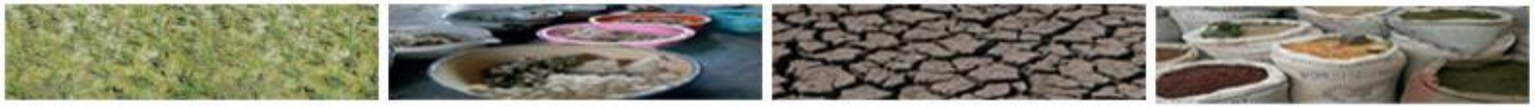
- **Interval**
  ex. Age 1 to n
- **Ratio**
  ex. Percentage of expenditure 0% to 100%

Categorical

- **Nominal**
  The categories are not ranked ex. 1=female, 2=male
- **Ordinal**
  The categories are ranked (ex. 1=poor, 2=medium, 3= good)

Type of variables

	Type of variables		
	Continuous	**Categorical**	
		Ordinal	***Nominal***
Do arithmetic operations on values make sense?	Yes	No	No
Are values ordered?	Yes	Yes	No
Types of values	Numeric	Alphanumeric codes	Alphanumeric codes

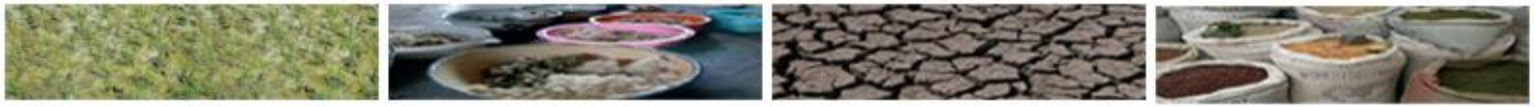


Continuous variables

Continuous variables assume numeric values which are expressed in a given unit of measurement

- Income, millimeters of rainfall, kilograms of agricultural production, percent of food insecure households, Weight-for-height z-score for children, etc.

Most importantly, each number (within a variable) has a meaning in relation to the other numbers, allowing arithmetic comparisons to be drawn



Categorical variables

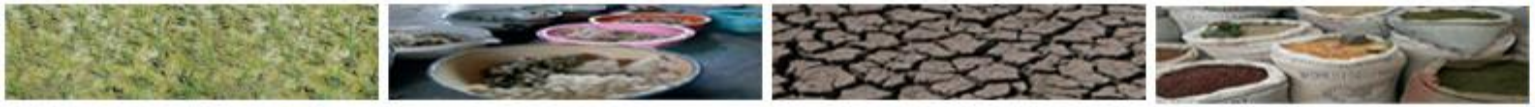
Categorical variables are used to describe the different types of properties the item of interest can have. Values are categories, taking a limited set of values.

Examples include:

- Sex of an individual = male or female
- Level of education = primary school, secondary school, university and above or no schooling
- Age group of a child = 0-11 months, 12-23 months, 24-35 months, 36-47 months, 48-59 months

Categories can be denoted numerically or alphabetically

Categorical variables take 2 forms-- ***nominal*** and ***ordinal*** variables.



Nominal variables

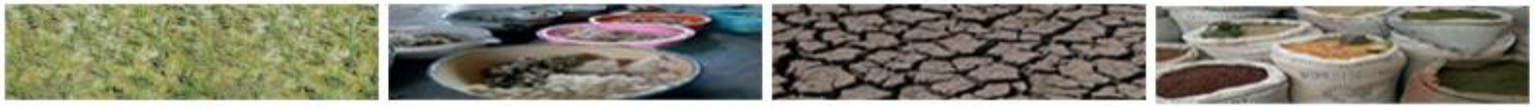
A nominal measurement scale is a set of mutually exclusive categories (only one category can be used to describe the subject of interest) that varies qualitatively but not quantitatively, for example gender, provinces, income sources, etc.

Codes are labels representing different behaviours/ characteristics and **they do not imply any underlying order.**

Ordinal categorical variables

An ordinal measurement scale differs from a nominal one in that the order among the original categories is preserved in the analysis. However differences between adjacent categories are not equal.

Examples social class and perception (bad – medium – good).



vam
food security analysis

Descriptive statistics

Descriptive statistics are the most basic from of statistics

They include:

- Summaries of one variable
- Comparisons of two or more variables

These tests are the foundation for more advanced statistical techniques




vam
food security analysis

Descriptive statistics

Continuous

- Range
- Mean
- Median
- Mode

Categorical

- Frequencies
- Crosstabs

vam
food security analysis

First lets discuss descriptives for continuous variables…




vam
food security analysis

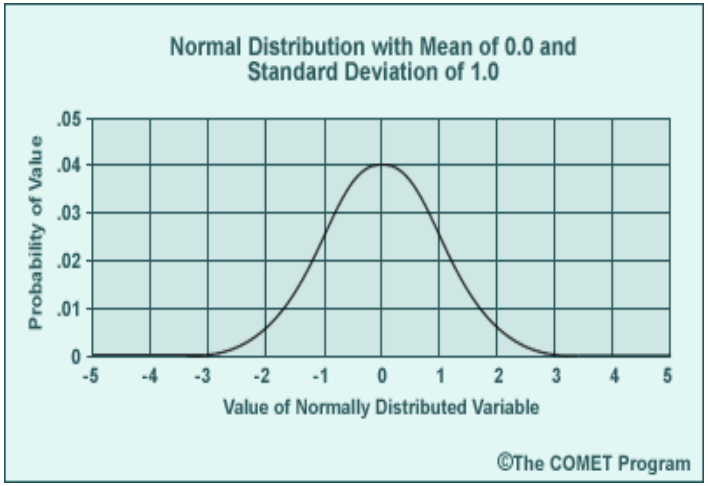
Range

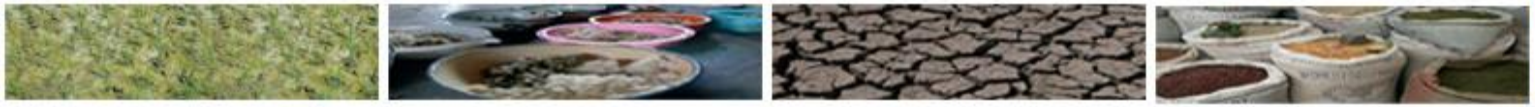
The range is the spread between the smallest and the largest values in a distribution

What is a mean?

The (arithmetic) **MEAN** is the sum of all the values divided by the numbers of cases

Statistics such as mean assume normal distributions





Median

The **MEDIAN** is the value above and below which half of the cases fall, the 50th percentile, i.e. the middle value of a set of observations ranked in order.

The median is a measure of central tendency not sensitive to outlying values--unlike the mean, which can be affected by a few extremely high or low values.

A median does not assume a normal distribution

Mode

The **MODE** of a distribution is the value of the observation occurring **most frequently**. It can be used with all measurement scales.

If several values share the greatest frequency of occurrence, each of them is a mode.




vam
food security analysis

To illustrate these concepts…

Looking at age data from 10 individuals…

1	2	3	4	5	6	7	8	9	10
12	19	23	26	28	28	28	34	36	38

- What is the range?
- What is the mean?
- What is the median?
- What is the mode?

vam
food security analysis

Now lets discuss descriptives for categorical variables…




vam
food security analysis

Analysing categorical data

If you want to look at the relationship between two categorical variables:

Mean and median not useful, except for when it is a binary variable (0/1)

Instead, we use frequencies and cross-tabs

Descriptives for categorical data

- The most basic descriptive for categorical variables is the **frequency**
  - which shows the number of cases
  - or % of cases

Sex of hh head

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Female	236	18.2	18.2	18.2
	Male	1064	81.8	81.8	100.0
	Total	1300	100.0	100.0	




vam
food security analysis

Descriptives for categorical data

- We can also cross-tabulate categories from one variable with categories from a second variable

	Continuous (scale)	**Categorical**
Continuous (scale)	none	Mean Median
Categorical	Mean Median	Frequency

Example: Scale and nominal

		Food Consumption Score
		Mean
Residence status	IDPs	64.0
	Residents	75.0
	Others	57.6
	Total	67.8

Example: Ordinal and nominal

		Food Consumption Groups			
		poor	borderline	acceptable	N. of cases
		Row N %	Row N %	Row N %	Count
Residence status	IDPs	7.1%	15.6%	77.3%	946
	Residents	2.2%	6.0%	91.8%	580
	Others	16.2%	20.2%	63.6%	52
	Total	5.6%	12.2%	82.2%	1578




vam food security analysis

Example… (row %)

		Food Consumption Groups			
		Poor	Borderline	Acceptable	Total
		Row N %	Row N %	Row N %	Row N %
Sex of hh head	Female	7.2%	14.8%	78.0%	100.0%
	Male	4.7%	9.0%	86.3%	100.0%
	Total	5.2%	10.1%	84.8%	100.0%

What percentage of female headed HHs have acceptable consumption?

What percentage of male headed HHs have poor food consumption?

What percentage of borderline consumption HHs in total?




vam food security analysis

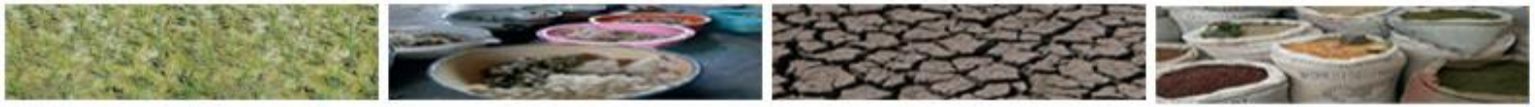
Example… (column %)

		Food Consumption Groups			
		Poor	Borderline	Acceptable	Total
		Column N %	Column N %	Column N %	Column N %
Sex of hh head	Female	25.4%	26.7%	16.7%	18.2%
	Male	74.6%	73.3%	83.3%	81.8%
	Total	100.0%	100.0%	100.0%	100.0%

What percentage of poor consumption HHs are female-headed?

What percentage of acceptable consumption HHs are male-headed?

What percentage of female-headed HHs in total?

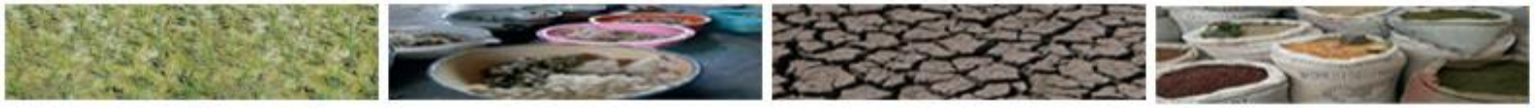


vam
food security analysis

DATA PREPARATION IN SPSS

Data view

- Cases and variables
  - **Case** → each row is one case
  - **Variable** → each column is one variable
  - The differences between the two can be easily seen by switching from data view to variable view in SPSS



vam food security analysis

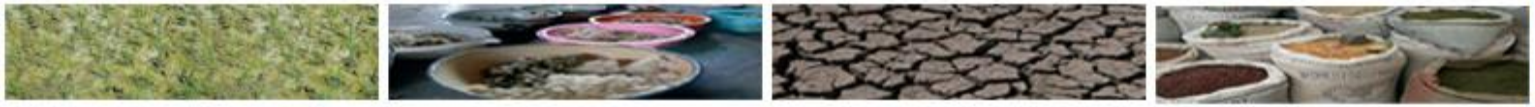
Data view

*Tajikistan FSMS dataset.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

58 : FSMS_ID | 1203052 | Visible: 456 of 456 Variables

	Month_Year	Round	HH_ID	FSMS_ID	s0_2	s0_3	Fewsnet_livelihood	zones	jamoats	s0_6	
1	Nov 2008	1	102011	1102011	9	03-Nov-2008	Southen Khatlon Cotton, Vegetabl...	Qumsangir	Panch	2	1 mai
2	Nov 2008	1	102012	1102012	9	03-Nov-2008	Southen Khatlon Cotton, Vegetabl...	Qumsangir	Panch	2	1 mai
3	Nov 2008	1	102013	1102013	9	03-Nov-2008	Southen Khatlon Cotton, Vegetabl...	Qumsangir	Panch	2	1 mai
4	Nov 2008	1	102014	1102014	9	03-Nov-2008	Southen Khatlon Cotton, Vegetabl...	Qumsangir	Panch	2	1 mai
5	Nov 2008	1	102015	1102015	9	03-Nov-2008	Southen Khatlon Cotton, Vegetabl...	Qumsangir	Panch	2	1 mai
6	Nov 2008	1	102016	1102016	9	03-Nov-2008	Southen Khatlon Cotton, Vegetabl...	Qumsangir	Panch	2	1 mai
7	Nov 2008	1	102017	1102017	9	03-Nov-2008	Southen Khatlon Cotton, Vegetabl...	Qumsangir	Panch	2	1 mai
8	Nov 2008	1	102021	1102021	9	04-Nov-2008	Southen Khatlon Cotton, Vegetabl...	Qumsangir	Panch	2	oktyab
9	Nov 2008	1	102022	1102022	9	04-Nov-2008	Southen Khatlon Cotton, Vegetabl...	Qumsangir	Panch	2	oktyab
10	Nov 2008	1	102023	1102023	9	04-Nov-2008	Southen Khatlon Cotton, Vegetabl...	Qumsangir	Panch	2	oktyab
11	Nov 2008	1	102024	1102024	9	04-Nov-2008	Southen Khatlon Cotton, Vegetabl...	Qumsangir	Panch	2	oktyab
12	Nov 2008	1	102025	1102025	9	04-Nov-2008	Southen Khatlon Cotton, Vegetabl...	Qumsangir	Panch	2	oktyab
13	Nov 2008	1	102026	1102026	9	04-Nov-2008	Southen Khatlon Cotton, Vegetabl...	Qumsangir	Panch	2	oktyab
14	Nov 2008	1	102027	1102027	9	04-Nov-2008	Southen Khatlon Cotton, Vegetabl...	Qumsangir	Panch	2	oktyab
15	Nov 2008	1	102031	1102031	9	05-Nov-2008	Southen Khatlon Cotton, Vegetabl...	Qumsangir	Panch	2	Rudak



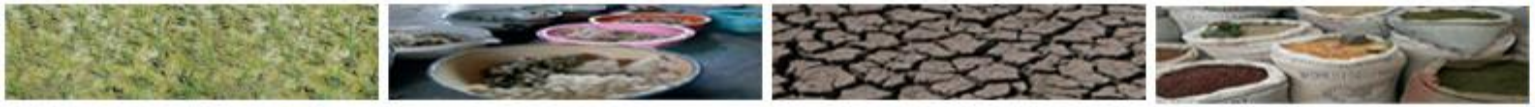
vam food security analysis

Variable view

*Tajikistan FSMS dataset.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

	Name	Type	Width	Decimals	Label	Values	Missing	Colu...	Align	Measure	Role
1	Month_Year	Numeric	8	0	Month and year	{1, Nov 200...	None	8	Right	Nominal	Input
2	Round	Numeric	8	0	Round number	None	None	6	Right	Nominal	Input
3	HH_ID	Numeric	7	0	Household code	None	None	7	Right	Scale	Input
4	FSMS_ID	Numeric	8	0	FSMS code (unique)	None	None	8	Right	Scale	Input
5	s0_2	Numeric	2	0	Interviewer team code	None	None	4	Right	Scale	Input
6	s0_3	Date	11	0	Date of interview	None	None	11	Right	Nominal	Input
7	Fewsnet_livelihood	Numeric	8	0	FEWSNET livelihood zones	{2, Western...	None	23	Right	Nominal	Input
8	zones	Numeric	2	0	Zone code	{1, Qumsan...	None	8	Right	Nominal	Input
9	jamoats	String	20	0	Jamoat name	None	None	9	Left	Nominal	Input
10	s0_6	Numeric	2	0	Jamoat code	None	None	3	Right	Scale	Input
11	village	String	20	0	Village name	None	None	12	Left	Nominal	Input
12	s0_8	Numeric	2	0	Village code	None	None	4	Right	Scale	Input
13	s0_9	Numeric	1	0	Household number	None	None	4	Right	Scale	Input
14	Adm1_name	Numeric	8	0	Admn1_name	{2847, Bada...	None	17	Right	Nominal	Input
15	s0_11	Numeric	1	0	Sex of hh head	{1, Female}...	None	5	Right	Nominal	Input



vam food security analysis

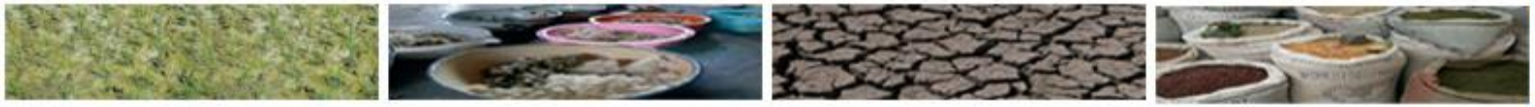
Prepare data before export (example EXCEL)

	B	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
1																			Q8 sou time
2	تاريخ المقابلة (اليوم/الشهر/السنة)	ID_final	Q1 Sex of the Family Head	Q2a Total Number of the persons in the Family	Q2b Total number of children under 18 years old:	Q2b Number of children less than 5 years (0-59 months) today:	Q3a How long the family has lived in Jordan?	Q3b Are you hosted by a resident family?	Q3 c If No (in 3b above), are you sharing with another Refugee family from Syria?	Q3d If yes (in 3b or 3c above), how many families are living here?	Q4a Where do you seek health assistance when sick ccurrently?(Ask the question and choose one number correspondin g to answer)	Q4b If 'No assistan ce' in Q4a, why?	Q5 Does the family have access to sufficient water for drinking, cooking, washing and toilet purposes?	Q6 What is the problem / major problems related to water and facing the family? (One answer or multiple answers)				Q7 Does the family have access to soap and Hygiene ?	First main source
3	08/10/2012	010601	m	5	3	3	2	1	0	6	4	0	1	0	0	2	1	1	0
4	08/10/2012	010602	m	4	2	0	2	0	0	0	6	0	1	0	0	2	1	1	0
5	08/10/2012	010603	m	4	2	2	3	1	0	0	4	0	0	0	0	0	3	0	0
6	08/10/2012	010604	f	5	4	1	2	1	0	2	5	0	1	0	0	2	1	0	0
7	08/10/2012	010605	m	6	4	1	2	0	0	3	4	0	1	0	0	2	3	1	0

vam food security analysis

Prepare data before export (example EXCEL)

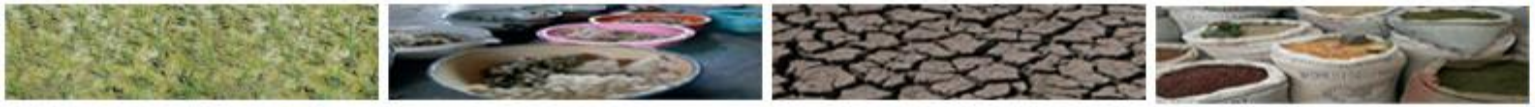
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	country	ID_final	Date	cluster_id	team_id	HH_id	Q1	Q2	Q2b_1	Q2b_2	Q3a	Q3b	Q3c
2	1	230302	15/10/2012	23	3	2	1	7	4	0	4	0	1
3	1	240305	14/10/2012	24	3	5	2	8	6	1	4	0	0
4	1	550411	23/10/2012	55	4	11	1	13	2	2	3	0	1
5	1	550412	23/10/2012	55	4	12	1	13	0	0	3	0	1
6	1	210411	13/10/2012	21	4	11	2	1	0	0	3	0	1
7	1	550102	23/10/2012	55	1	2	1	6	6	3	3	1	0
8	1	200407	13/10/2012	20	4	7		4	2	0	3	0	1
9	1	40505	08/10/2012	4	5	5	1	1	0	0	3	0	0
10	1	250306	15/10/2012	25	3	6	2	6	2	2	3	0	0
11	1	20303	08/10/2012	2	3	3	1	3	1	0	3	1	0
12	1	410102	15/10/2012	41	1	2	1	12	4	1	3	0	0
13	1	120303	10/10/2012	12	3	3	1	1	0	0	3	0	1
14	1	450214	18/10/2012	45	2	14	1	6	4	2	3	0	0
15	1	30506	08/10/2012	3	5	6	1	8	5	0	3	0	0
16	1	130401	08/10/2012	13	4	1	1	6	3	1	3	0	0



vam
food security analysis

Importing data in SPSS:

- Open SPSS is as simple as selecting
- → File → Open → Data
- Select files by type
- Browse and select data
- Open

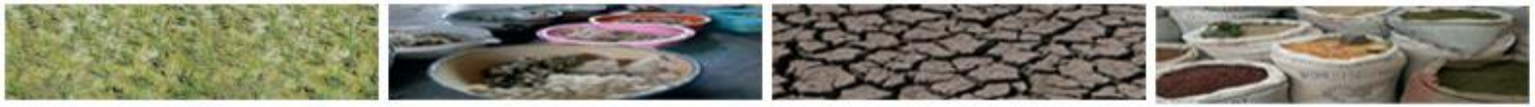


Clean variable view

- **Name**: short name, input variables usually refers to section in questionnaire (e.g. s2_3a, s2_3b); must be unique
- **Type**: number="numeric", text="string", date="date"
- **Width**: Maximum number of characters allowed
- **Decimals**: Number of decimals (e.g. 2=0.23)
- **Labels:** description of the variable
- **Values**: Codes for categorical (nominal) variables need to be defined
- **Missing values**: Values that should be excluded during the analysis (e.g. 99= not applicable, 88=not known, etc.)
- **Width of column**: Width how column should appear on the screen
- **Alignment**: If text = "left", if number = "right"
- **Level of measurement**: if number that can be divided = "scale", if categorical number or string = "nominal" (e.g. 1=yes, 0=no) , if rank = "ordinal" (e.g. 1=poor, 2=medium, 3= better-off)

Data cleaning

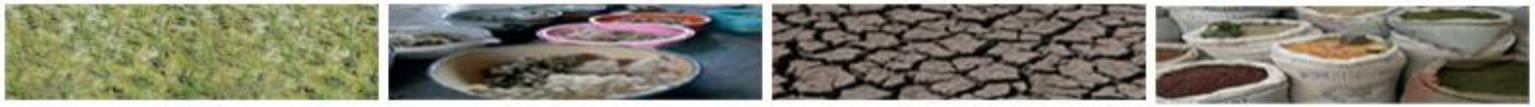
- Cleaning data can be a painful process
- Being systematic about cleaning data from the beginning of the process can save hours of work later in the analysis
- Some basic tools to use in SPSS to clean data:
  - Sorting cases in ascending or descending order
  - Run quick frequency tables for categorical variables
  - Identify duplicate cases – shows cases which have the same unique identifier



vam
food security analysis

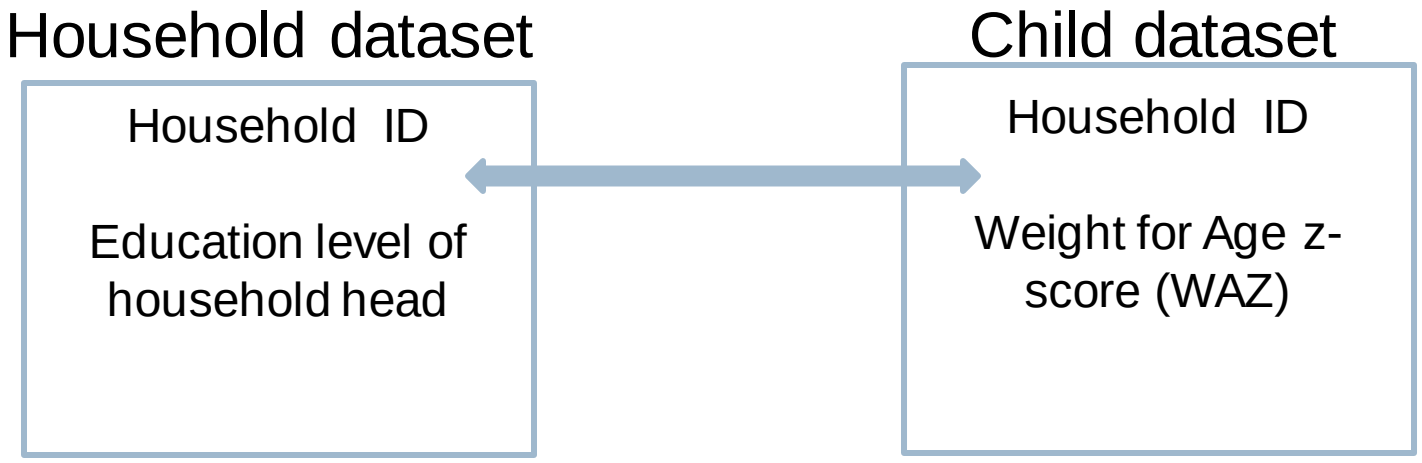
Merging datasets

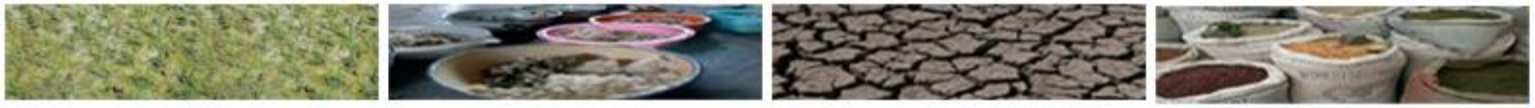
- For each level of analysis, there is one SPSS dataset (e.g. HH-level, child-level, HH-member level, market price data);
- It is possible to link data from various levels by merging datasets using the **adding variables** function;
- It is also possible to combine datasets that collected the same indicators (e.g. data from different countries, different rounds) using the **adding cases** function;



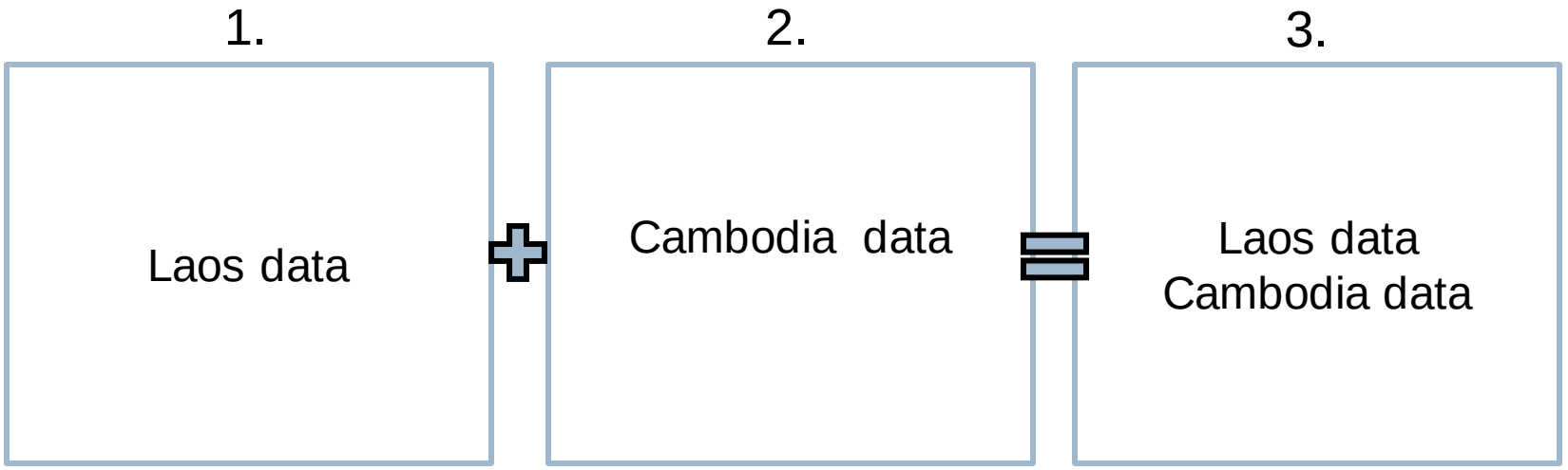
Merging datasets (adding variables)

- For example – the education level of a household head is recorded in the household dataset. We may be interested to find if the nutritional status of a child is related to education of the household head. But the child data is in a separate dataset.
- In order to merge the datasets, a common variable must exist in each dataset. In this case, a household identifier must be in both datasets.





Merging datasets (adding cases)



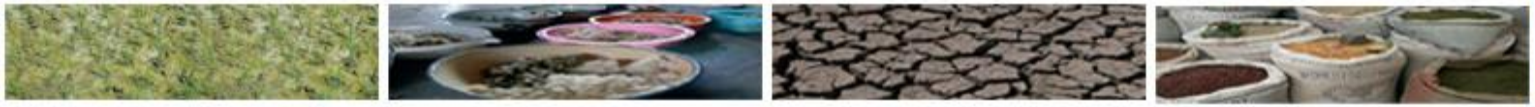
Create unique codes

- First column: round number (01-…..)
- Second column: region code (01-10)
- Third column: site code (01-25)
- Fourth column: HH number (01-15)

- How many digits will the unique code have?

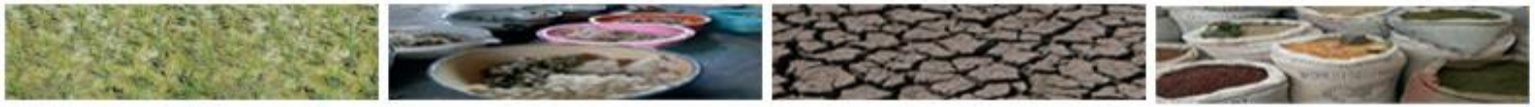
- _ _(round)_ _(region) _ _(site-code) _ _(HH-number)

Questions?



vam
food security analysis

SPSS COMMANDS FOR DESCRIPTIVE ANALYSIS



vam
food security analysis

SPSS commands

- Recode into different variable
- Compute
- Filter
- Custom tables
- Paste syntax




vam food security analysis

Create new variables using recode

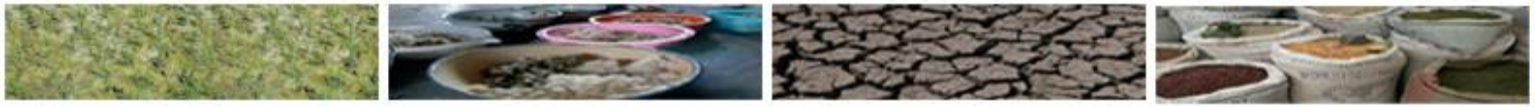
- Recoding a variable is the most common command we use in SPSS
  - Into same variable
  - Into different variable
- Objective: Classify categorical or scale variables into groups
  - Drinking water sources into “improved”/”not improved”
  - Age into age groups
  - FCS into FCS groups

4.2 What is the main source of drinking water for your household?

(Circle one)

- 1 = Piped water
- 2 = Well (protected)
- 3 = Well (unprotected)
- 4 = River, stream or pond
- 5 = Collecting rainwater
- 6 = Tanker truck water

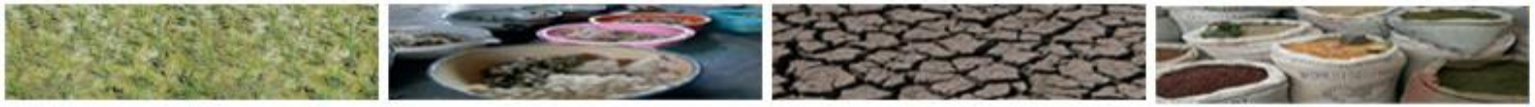
Improved source	Unimproved source
Piped water	Well (unprotected)
Well (protected)	River, stream or pond
Collecting rainwater	Tanker truck water



Creating a new variable using compute

- Computing a new variable is usually done when a mathematical formula is used to derive a new variable
  - E.g calculating total HH size by adding age groups
  - % Female in each HH
  - % of dependent HH members in each HH (dependent <15, 60 plus)

1.5 Please complete the household demographics table on the right. Record the number of individuals in each age category, differentiated by males and females.	Age	Male	Female
	a. 0-5 years	___	___
	b. 6-14 years	___	___
	c. 15-59 years	___	___
	d. 60 years or older	___	___



vam
food security analysis

Selecting cases using filters

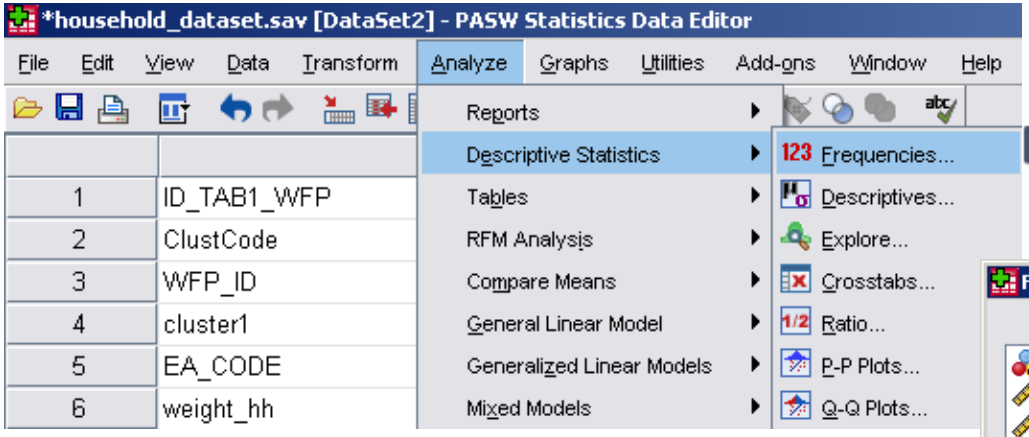
- Often, analysis or data processing should only be applied to a subset of the cases in your data
- For example, we might want to run analysis only on IDPs
- This is done easily in SPSS using the SELECT CASES function
- Select Data > Select Cases and then use the If conditional dialog box in the subsequent screen in SPSS to write a condition. In this case: select cases if residence status = 1 (assuming IDPs = 1)

Selecting cases using filters

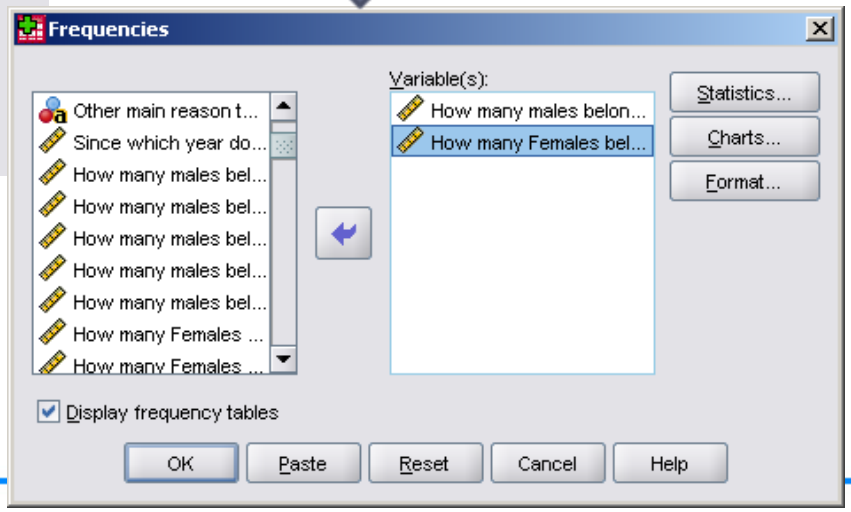
- SPSS will create a filter variable at the end of the dataset and only select cases that meet the criteria you've provided
- Ensure to select all cases from the same menu once you've finished the analysis on the subset

Frequencies

- Select ANALYZE > Descriptive Statistics > Frequencies

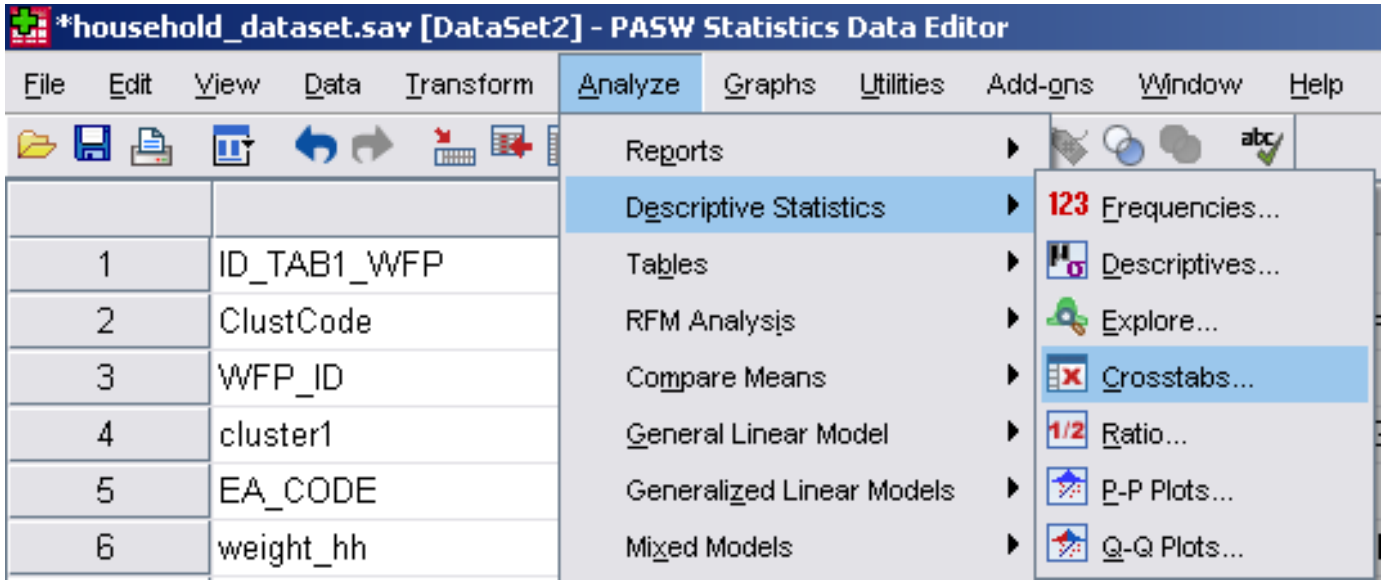


Select the variable you want to see

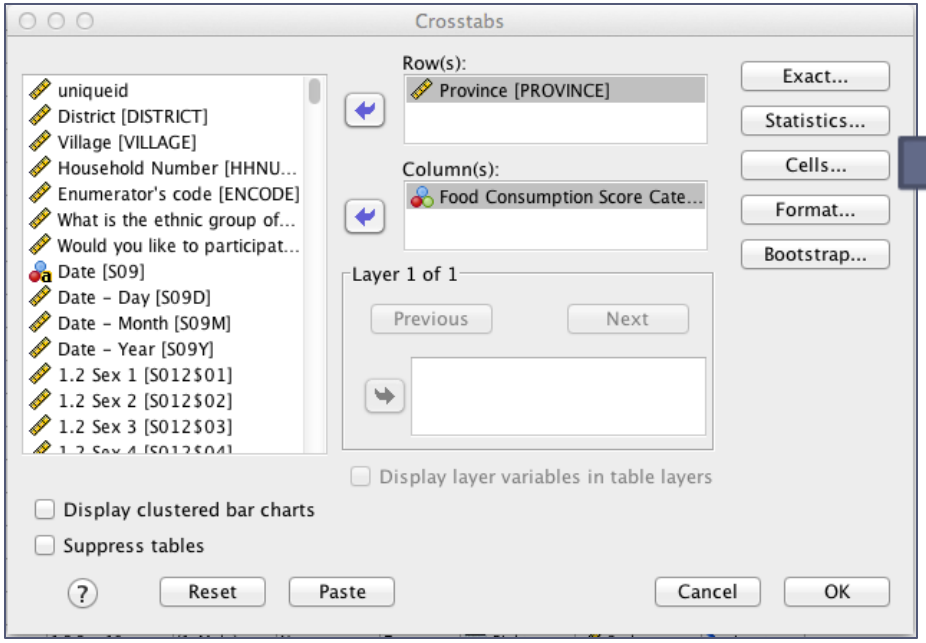


Crosstabs

- Select ANALYZE > Descriptive Statistics > Crosstabs

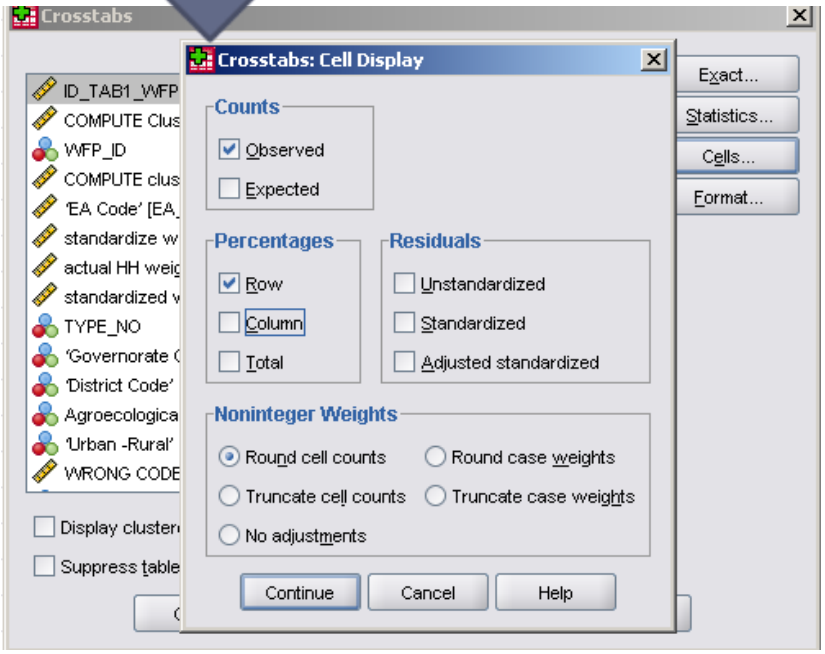


Crosstabs



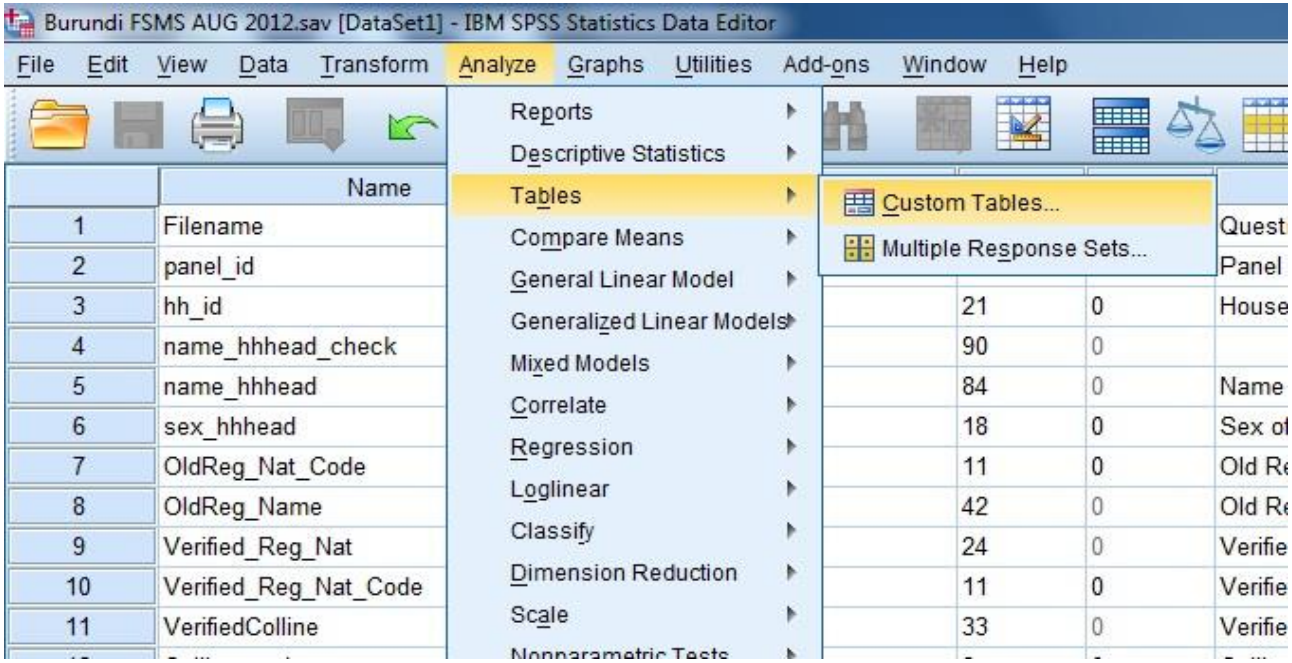
Select the variables you want to cross-tabulate

Click on “Cells” to view by percentages

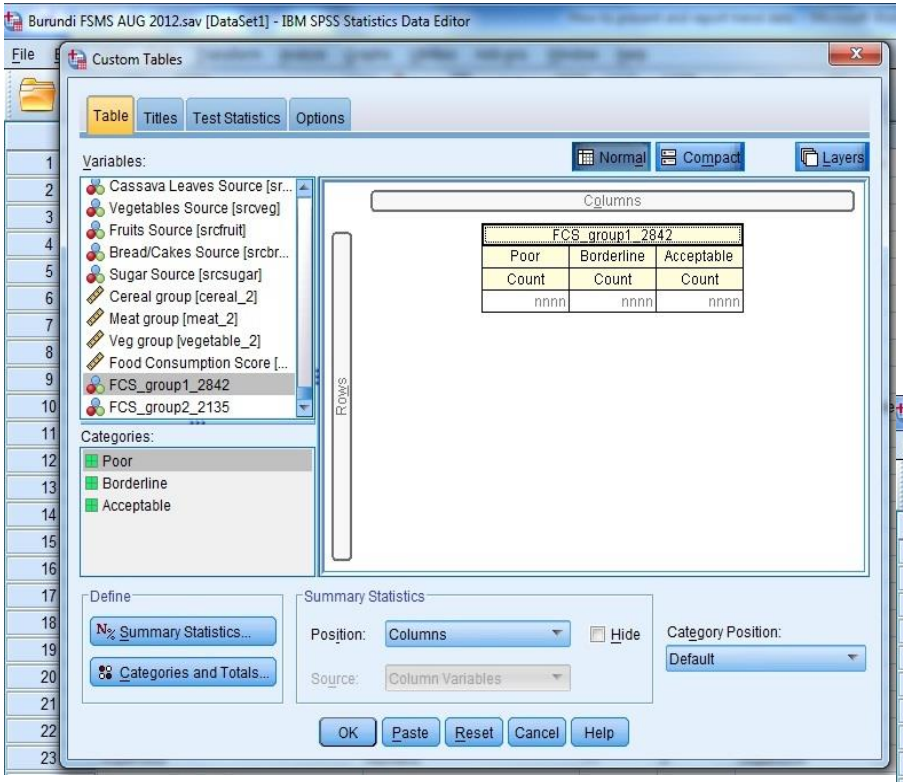


Custom tables

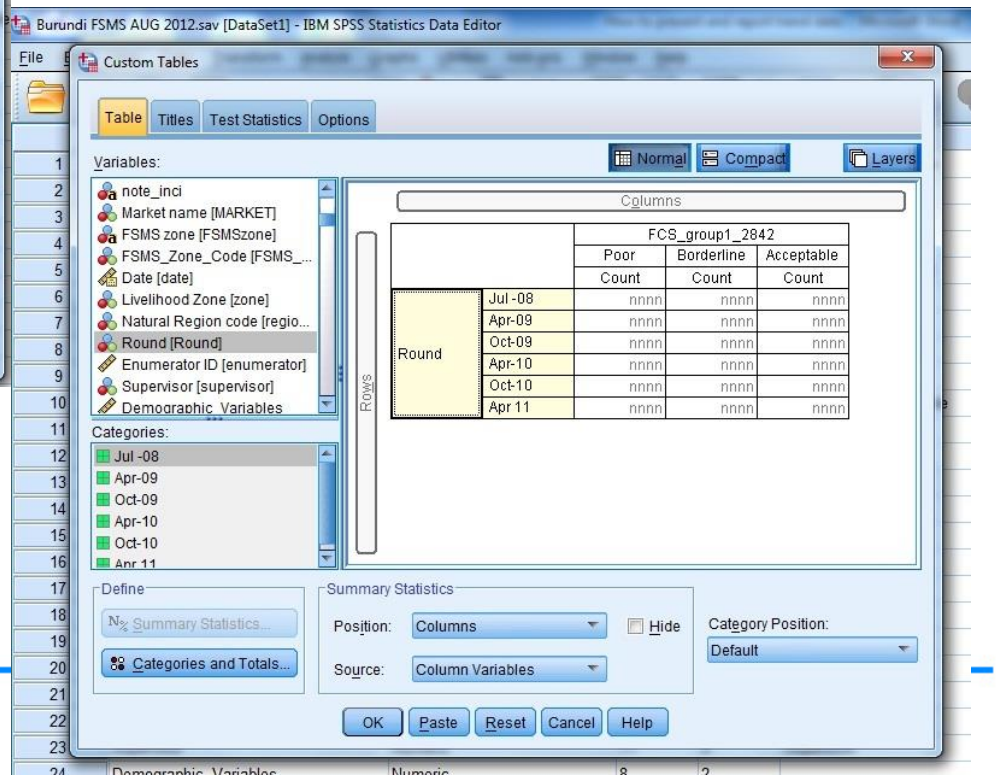
- Select > ANALYZE > select TABLES > select CUSTOM TABLES



Custom tables

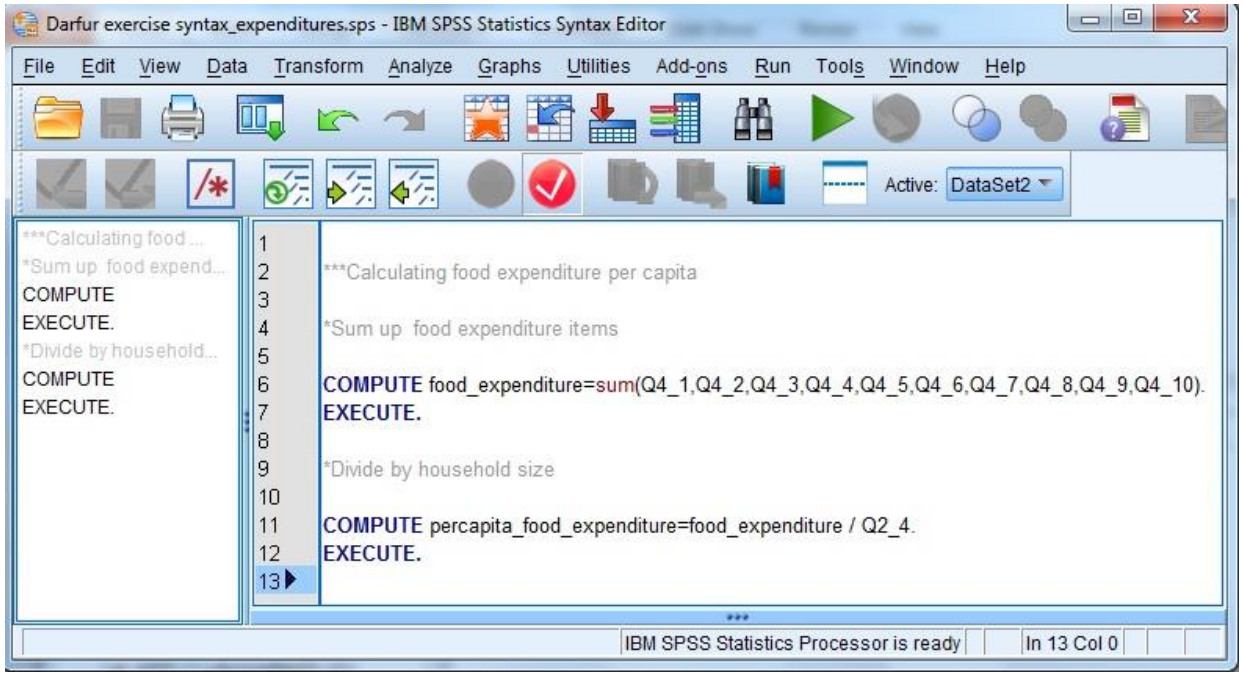


Select the variable you want to monitor in the variable list and drag and drop it into the columns & rows area on the canvas pane



Paste syntax

- We want to repeat steps, it is useful to create syntax



Types of variables

	Continuous	Categorical	
SPSS terminologies	Scale	Ordinal	Nominal
Which mathematical function can be applied?	- Calculations ("compute") - Categorizing ("recode into different variable")	Categorizing ("recode into different variable")	Categorizing ("recode into different variable")
Is there an order?	Yes	Yes	No
Types of values	Numbers or percentages	Codes (order has meaning)	Codes (order has no meaning)