VAM Guidance Paper

Creation of a Wealth Index





Prepared by: Lisa Hjelm, Astrid Mathiassen, Darryl Miller, Amit Wadhwa.

Reviewed by: Rossella Bottone (June 2017)

Contact: Claudia AhPoe (<u>claudia.ahpoe@wfp.org</u>) or Kurt Burja (<u>kurt.burja@wfp.org</u>)

Contents

Background	4
Measurements of wealth	4
Current use of wealth index (WI)	4
Use of the wealth index in WFP's household surveys	4
Urban-rural considerations	5
Steps for the creation of a wealth index (WI)	6
STEP 1: Select variables	6
STEP 2: Explore variables	7
Step 3: Recode into scale variables	11
Step 4: Principal Component Analysis (PCA)	15
Step 5: Create wealth index quintiles	22
Step 6. Graph the wealth index	24
Step 7. Select the final result and report the variables	25
Additional reading:	26

Background

Wealth is the value of all natural, physical and financial assets owned by a household, reduced by its liabilities. Household wealth is a measure commonly used in food security assessments. It gives an idea of households' ability to access food, the severity of food insecurity and gives information about the economic situation of the food insecure. It is used to differentiate between the poorer and the wealthier households in food security related indicators, such as food consumption, and thereby provides information on how to target the food insecure.

This guidance aims at complementing the Comprehensive Food Security & Vulnerability Analysis Guidelines (CFSVA 2009) with a practical step-by-step guidance on how to create one, while providing VAM officers and other food security analysts with a more general background of the wealth index (WI).

Measurements of wealth

There are several ways in which wealth, economic status of households and living standards can be measured. Income, expenditure and consumption are three common measurements.

However, there are challenges in collecting and measuring income and expenditure accurately. An alternative is to use data on asset ownership and housing characteristics and combine this information into a proxy indicator such as the wealth index, which is created using principal component analysis (PCA). Asset ownership gives an indication of the longer-term economic status of a household and is less dependent on short-term economic changes compared with other wealth or poverty measures.

The wealth index measures relative wealth and, unlike a poverty line, is not an absolute measure of poverty or wealth. When referring to the wealth of households based on the wealth index we can talk about poorer and wealthier households but we cannot conclude who is absolutely poor and wealthy. The wealth index quintiles divide the whole population into five equally large groups, based on their wealth rank. For example, in an area where only 10% of households fall below the poverty line, 40% of households will still fall into the two poorest quintiles and therefore be classified as the poorest.

Current use of wealth index (WI)

The wealth index is commonly used in reports and analysis based on datasets from Demographic and Health Surveys (DHS), Multiple Indicator Cluster Surveys (MICS), and WFP surveys (mainly CFSVAs¹ and EFSAs), and is used to rank households into quintiles. The value of using the wealth index is especially recognised in contexts where reliable income and expenditure data is absent.

The research questions related to the wealth index vary according to the different interests of the surveys. In DHS, the wealth index is chosen because of the major impact that wealth status has on household level health. It allows the researchers to identify the impact of wealth status on health outcomes². For MICS, the wealth index serves a similar purpose in terms of understanding health outcomes. It is also used to target poverty alleviation programmes and projects³.

Use of the wealth index in WFP's household surveys

For WFP-VAM food security analyses, the wealth index quintiles are useful for cross-tabulation with other variables in the dataset. For example, cross-tabulating with regions can show areas with higher proportions of poor households or cross-tabulating with food consumption groups could determine

¹ WFP, Comprehensive Food Security & Vulnerability Analysis Guidelines, 2009 page 211

² DHS wealth index. See at. http://dhsprogram.com/topics/Wealth-Index.cfm

³ UNICEF(2008). Measuring child poverty. See at

http://www.southampton.ac.uk/ghp3/docs/unicef/presentation2.3.pdf

what proportion of households with poor food consumption are also in the poorer groups. In addition, the wealth index can be used as a proxy for food access.

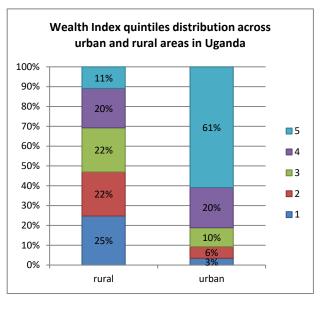
Urban-rural considerations

One consideration that should be taken into account in relation to the wealth index is that wealth is characterized by ownership of different types of assets in urban areas compared with rural areas. Depending on the variables included in the index, the wealth measure can be biased towards urban or rural households. One solution is to include variables that are valid as proxies of wealth in both urban and rural areas. For example, if a high percentage of households live in urban areas and few households practise agriculture, we may consider excluding productive assets and livestock.

If the living conditions in urban and rural areas are very different, another approach can be to create separate indices for urban and rural areas. The following chart shows the distribution of rural and urban households in Uganda using the uniform national wealth index. We observe a remarkable urban-rural divergence in terms of relative wealth status. The analysts may need to assess if this distribution reflects the subsutantial urban-rural inequality or if it is a result of variable selection

bias. If the latter situation is likely to be true, we may need to construct a separate wealth index for rural and urban households that take into account the differences in assets owned.

The question of when to create a separate urban-rural wealth index is an open one that calls for both researchers and analysts to examine the specific context. In this guidance, we focus on the methods of constructing a national wealth index. Similar methodologies can be adopted for index creation separately for urban and rural areas.



Steps for the creation of a wealth index (WI)

- 1. Select variables
- 2. Explore variables
 - a. Frequencies
 - b. Missing values
- 3. Recode into binary variables
- 4. Principal components analysis (PCA)
- 5. Create wealth index quintiles
- 6. Graph the index
- 7. Select the final result and report the variables

Note: Uganda LSMS 08/09 dataset is used to demonstrate the WI creation and SPSS (Statistical Package for the Social Sciences) procedures in this guidance.

STEP 1: Select variables

The first step in creating the wealth index is to select variables to be incorporated in the questionnaire. The questionnaires ask what the households own based on an extensive list of assets and other housing characteristics that reflect their economic status (see example in the table below).

A broad range of variables could be included in the analysis: a greater number can reduce the sampling bias and generate a better distribution of households. The final list should be country-specific, and simultaneously capture the differences in ownership among households (see more in Step 2).

Productive assets	Non-productive assets	Household utilities and other
Hand mill	Radio	Types of:
Sickle	Refrigerator	Water supply
Axe	TV	Toilet
Livestock	Bicycle	Flooring
Ное	Motorbike	Walls/house
Tractor	Phone/cell phone	Roof
Plough	Chair	Light source
	Table	
	Bed	Persons sleeping per room
		Land ownership
		Livestock ownership

Example of variables that can be included in household surveys:

STEP 2: Explore variables

Once the possible variables have been identified in the data set, they require further investigation before being selected for the index.

2.a. Explore the variables by running descriptive analysis including a frequency of each variable.

As a first step in exploring the variables to include in the index some basic cleaning of data may be needed. A household which has missing values for any of the assets will be excluded in the wealth index construction. If a substantial proportion of missing values is detected, the analysts should check the data quality again and if possible go back to the enumerators to ensure accurate data collection and entry.

We need to select the variables that are capable of distinguishing relatively "wealthy" households and relatively "poor" ones. The rule of thumb is that if a variable/asset is owned by more than 95% or less than 5% of the sample, it should be excluded from the analysis. For example, knowing that 99.2% of Ugandan households don't own a generator will not help the analyst to distinguish between richer and poorer households by this asset ownership (see table below). Thus, this variable will be excluded from the index.

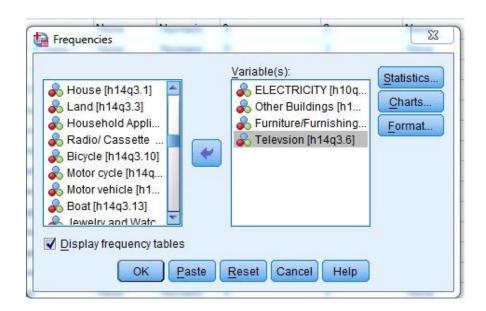
		Frequency	Percent	Valid Percent	Cumulative
					Percent
	no	5152193	98.6	99.2	99.2
Valid	yes	44014	.8	.8	100.0
	Total	5196207	99.5	100.0	
Missing	System	28473	.5		
Total		5224680	100.0		

DOES THIS HOUSE HAVE A GENERATOR? (SPSS frequency table)

ile <u>E</u> o	dit	<u>V</u> iew <u>D</u> ata	<u>T</u> ransform	Analyz	e <u>G</u> raphs	s <u>U</u> tilitie	s Ado	d- <u>o</u> ns	Windo						
	H			R	e <u>p</u> orts		•	HA						2	
				D	escriptive \$	Statistics	- F	123	<u>F</u> requer	icies	-				
-		Name	Label	Ta	a <u>b</u> les		•	μ۹	<u>D</u> escrip	tives		Missing	Columns		Align
1		hhid	unique hh	С	o <u>m</u> pare Me	eans	•	<u>_</u>	Explore.			None	20	IIII Ri	-
2		very_high_fo		G	eneral Line	ear Model	•		— · Crossta	hs		None	8	III Ri	-
3		high_foodexp	=1 if more	G	enerali <u>z</u> ed	Linear Mo	dels		Ratio			None	8	III Ri	•
4		fcs_poor_bo		М	i <u>x</u> ed Model	s	•		_			None	8	III Ri	-
5		veryhigh_st	=1 if more	<u>c</u>	orrelate		•		P-P Plot			None	8	III Ri	-
6		food_dep	=1 if food e	R	egression		•	₹,7	Q-Q Plo			None	8	를 Ri	-
7		fcs_2groups	RECODE	L	oglinear		•			2		None	8	IIII Ri	•
8		WI_quint	Wealth ind	С	lassify)		None	5	IIII Ri	•
9		exp_quintiles	constant_p	D	imension I	Reduction	•)		None	10	IIII Ri	-
10		crosswgt09	Weights w	s	cale		•			2		None	8	를 Ri	-
11		mult	panel weig		onparame	tric Tests				2		None	8	III Ri	•
12		pnlwgt09	panel weig	_	orecasting					2		None	8	IIII Ri	•
13		region	region		urvival					2		None	8	I Ri	-
14		urban	urban/rural	_			, r	<u> </u>)		None	8	I Ri	
15		regurb	regionxurb		ultiple Res)		None	8	I Ri	
16	_	district	District co		omp <u>l</u> ex Sa)		None	8	I Ri	
17		h1bq2a	Day		uality Cont		•)		None	8	I Ri	
18		month	Month		OC Curve)		None	8	III Ri	
19		h1bq2c	Year	No		lumeric	8)		None	8	III Ri	•
20	_	hsize	Number of	•		lumeric	9)		None	8	를 Ri	•
21	_	h9q03	HOW MAN				-)		None	8	를 Ri	•
22		crowdina 1	hh membe	r No	ne N	lumeric	8			>		Vone	10	I Ri	aht 👔
Data Vie		Variable View													

SPSS: Analyse \rightarrow Descriptive statistics \rightarrow Frequencies

In the dialogue box, click the variables in the left field that we want to run frequencies for and click on the right arrow to move the variables into the right field called **'Variable(s)**'. The option **'Display frequency tables**' at the bottom of the dialogue box is checked as default.



2 b. Run the frequencies for urban and rural areas separately

In a similar way to the rule of thumb discussed in 2.a, we run the frequencies for urban and rural areas separately to determine the variables to create a national wealth index. If there are certain assets owned by very few in either urban or rural areas, we will consider not including them because the national index needs to represent both urban and rural households. The inclusion of productive assets/livestock and land ownership should be reassessed if a high percentage of households do not practise agriculture or if many households are located in urban areas.

SPSS: Data→ Select cases

ta *ugano	da_fsindicat	tors for W	/I_28 nov.sav	[DataSet1]	- IBM SPS	S Statistics	Data Editor		
<u>F</u> ile <u>E</u> d	lit <u>V</u> iew	<u>D</u> ata	<u>T</u> ransform	<u>A</u> nalyze	<u>G</u> raphs	<u>U</u> tilities	Add- <u>o</u> ns	<u>W</u> indow	/ <u>H</u> elp
			efine <u>V</u> ariable et Measurem		h				
1	hhid		opy Data Pro				Width 10	2	Decima
2	very_	· ····	ew Custom A	-			Э	2	
3	high_		efine Dates		Cata		9	2	
4	fcs_p veryh		efine <u>M</u> ultiple				9	2	
6	food		entify D <u>u</u> plica	ile Cases	•		9	2	
7	fcs_2		ort Cases ort Variables				19	2	
8	WI_q		anspose				3	0	
9	exp_	м	erge Files				3	0	
10 11	Cross	1	estructure				9	2	
12	mult pnlwo	Ag	gregate				9	2	
13	regio		opy <u>D</u> ataset				8	2	
14	urbar	Sp	Split File 3 Select Cases 3 Weight Cases 10					0	
15	regur	<u><u> </u></u>						0	
16	distri	4 4 W						0	
17 18	h1bq mont		Month	None	Nu		B B	0	
10	h1bq		Year	None			B	0	

When the dialogue box pops up, select the urban/rural categorical variable ('urban/rural identifier' in this database) in the left field and select the option '**If condition is satisfied**' in the right field. Once selected, click the button '**If'** to customize specific conditions for the cases we want to analyse.

Image: Select Image: Select <td< th=""><th> unique hh identifi = 1 if more than 7 = 1 if more than 6 fcs_poor_border = 1 if more than 7 = 1 if food energy RECODE of fcs Wealth index qui constant_pc_exp Weights without panel weight UN region [region] urban/rural identifi Day [h1bq2a] Month [month] Year [h1bq2c] Number of neonl </th></td<>	 unique hh identifi = 1 if more than 7 = 1 if more than 6 fcs_poor_border = 1 if more than 7 = 1 if food energy RECODE of fcs Wealth index qui constant_pc_exp Weights without panel weight UN region [region] urban/rural identifi Day [h1bq2a] Month [month] Year [h1bq2c] Number of neonl
---	---

In this step, we want to look at data for urban and rural areas separately. Double click the selected categorical variable ('urban/rural identifier' in this database) and type =1' (in this dataset, urban=1 and rural=0) in the blank column right of the arrow. Now we are narrowing the analysis scope down to urban households.

Select Cases: If	artises of the second second second	X
 =1 if food energy RECODE of fcs Wealth index qui constant_pc_exp Weights without panel weight UN region [region] urban/urual identif regionxurban - su District code [dist Day [11bq2a] Month [month] Year [h1bq2a] Number of peopl HOW MANY ROO H members per improved constru improved constru improved constru improved constru improved constru 	✔ Urban=1 + 7 8 9 - 4 5 6 • 1 2 3 / 0 . ** () Delete ✓	Function group: All Arithmetic CDF & Noncentral CDF Conversion Current Date/Time Date Arithmetic Date Creation
	Continue Cancel Help	

Click **'Continue**' and repeat the 'frequency' process in 2.a. to examine the asset ownership frequencies of urban households. To analyse rural households, change the 'if' conditions into 'urban/rural identifier=0' and follow the same procedures as above.

After steps above, remember to turn off the selection by going to **Data \rightarrow **Select cases** and click **`All cases**'.

Step 3: Recode into scale variables

Before creating the wealth index, all variables should be transformed into scale variables. Most often asset ownership questions are categorical variables. Yes/no variables should be recoded into binary variables (variables that takes only 0 and 1 to represent the categorical effect) and variables with more than 2 categories should be transformed into bivariate variables.

Questions regarding housing characteristics and access to services are commonly categorical variables with several options. When this is the case a decision has to be made on how to recode these variables so that there are only two categories. When doing this, identify the alternatives that are more likely to be found in wealthy households compared with poorer households. How the variables should be recoded depends on the context of the country and what is more likely to distinguish poorer households from wealthier households. For instance, if the light source has many options, such as "none," "wood fire," "oil lamp," "petrol light," "electricity," it might be appropriate to recode this into "none/primitive" for those households that answered "none" or "wood fire" and "purchased energy source," for the remaining light sources. In another country, the same options might be recoded differently: "no electric" versus "electric."

The choice is based on what is more likely to define wealth and also by looking at the prevalence of both categories: if the prevalence is between 30 and 70%, then the indicator will probably help categorize more households than if the prevalence is only 5%. Two similar variables can be combined if this will result in a summed prevalence of between 30 and 70%.

For sanitation facilities and source of water the UNICEF/WHO standards can be used (see table below)⁴. However, the recoding between improved/ not improved is just one possibility. Other classifications can be used if this is more likely to separate the poorer households from the richer in the country context. One example is the alternative classification of source of water: bottled water is regarded as an unimproved source since quantities are not usually large enough to supply a household, but in reality, those who can afford to buy bottled water, especially in less developed countries, are often wealthier. So we may consider including bottled water in the improved group.

⁴ UNICEF&WHO. Progress on sanitation and drinking water.2013 Update. See at http://apps.who.int/iris/bitstream/10665/81245/1/9789241505390_eng.pdf

UNICEF/WHO Types of	UNICEF/WHO Types of drinking water sources ⁵				
IMPROVED	UNIMPROVED				
Drinking water sources	Drinking water sources				
Piped water on premises	Surface water +				
Public taps or standpipes	Unprotected dug well				
Tube wells or boreholes	Unprotected spring				
Protected dug wells	Cart with small tank/drum				
Protected springs	Surface water				
Rainwater collection	Bottled water				
Sanitation facilities	Sanitation facilities				
Flush/pour flush to:	Open defecation				
Piped sewer system	Shared sanitation facilities*				
Specific tank	Pit latrines without a slab or platform				
Pit latrine	Hanging latrines				
Ventilated improved [VIP] pit	Bucket latrines				
Pit latrine with slab					
Composting toilet					
+ Surface drinking water sources include river, dam, lake, pond, stream, canal, irrigation channels.					
* Sanitation facilities of an otherwise acceptable type shared between two or more households are shared sanitation facilities. Shared facilities include public toilets.					

When the variables have been reclassified, we assign categories values with 0 and 1. It is important to keep a record of how the variables have been recoded which the analyst can track and refer to during the analysis process. The record can also help the analyst adjust the recoding accordingly over time.

Name of the Variable	Wealthier	Poorer
Material of the house	1 = concrete or wood	0 = mud or thatch
Roof material	1= tiles or galvanized iron or concrete	0 = mud or thatch or plastic
Crowding	1 = 5 or fewer people per room	0 = 6 or more people per room
Type of lighting	1 = electricity or gas	0 = candle or wood
Source of water	1 = piped into dwelling or borehole with pump or protected dug well	0 = pond or unprotected well
Toilet facilities	1 = flush or ventilated improved latrine	0 = open pit or none (bush field)
Has a sewing machine	1 = yes	0 = no
Has a radio	1 = yes	0 = no
Has a TV	1 = yes	0 = no
Has a stove	1 = yes	0 = no
Has a fridge	1 = yes	0 = no
Has a mobile phone	1 = yes	0 = no
Has a bicycle	1 = yes	0 = no
Has a motorbike	1 = yes	0 = no
Has a car	1 = yes	0 = no

SPSS: Transform → Recode into same variables

ta *uganda_f	a *uganda_fsindicators for WI_28 nov.sav [DataSet1] - IBM SPSS Statistics Data Editor					
<u>F</u> ile <u>E</u> dit	<u>V</u> iew <u>D</u> ata	<u>Transform</u> <u>A</u> nalyze <u>G</u> raphs <u>U</u> tilit	ies Add- <u>o</u> ns <u>W</u> indo			
		Compute Variable Count Values within Cases				
42	Name	Shi <u>f</u> t Values	abel			
13	region	🔤 Recode into Same Variables				
14	urban	Recode into Different Variables				
15	regurb					
16	district	🕍 Automatic Recode				
17	h1bq2a	📲 Visual <u>B</u> inning				
18	month	Ran <u>k</u> Cases				
19	h1bq2c	🛱 Date and Time Wizard				
20	hsize					
21	h9q03	Create Time Series	OLD OCCI			
22	crowding	Replace Missing <u>V</u> alues				
23	improved_roo	Random Number Generators				
24	improved_w	Run Pending Transforms	Ctrl+G			
25	improved_floo	r improved construction material of flo	or			
26	improved_w	improved source of water				
27	improved_s	improved toilet facility				
28	h10q1	ELECTRICITY				

When the dialogue box pops up, select the variable ('crowding' which indicates the number of household members per room in this example) we want to recode from the left field and click the

arrow to move it into the right field. Then click the button '**Old and New Values**' to set the new values.

ţ	Recode into Same Variables
	 Recode into Same Variables regionxurban - subregions [regurb] District code [district] Day [h1bq2a] Month [month] Year [h1bq2c] Number of people in household [hsize] HOW MANY ROOMS DOES YOUR HOUS improved construction material of roof [im improved construction material of floor [i
	OK Paste Reset Cancel Help

According to the binary classification for 'crowding' (see chart on P10), click '**Range**' in the '**Old Value**' field and enter '0' through '5'. Then click '**Value**' in the '**New Value**' field and enter '0'. Now we recode the 'number of household members per room \leq 5' into '0'. Click the button '**Add**' to record this recoding.

Recode into Same Variables: Old and New Values	X
Old Value O Value: O value:	New Value © Value: 0 © System-missing
System-missing System- or user-missing Range: 1 1 5 0 Range, LOWEST through value:	Ol <u>d</u> → New: Change Remove
© Rang <u>e</u> , value through HIGHEST: O All <u>o</u> ther values	Cancel Help

-Old Value	New Value
© <u>V</u> alue:	Value:
	◎ System-missing
◎ <u>S</u> ystem-missing	
◎ System- or <u>u</u> ser-missing	Ol <u>d</u> > New: 0 thru 5> 0
Range:	0 unu 5> 0
	Add
through	Change
Range, LOWEST through value:	Remove
Range, value through HIGHEST:	
O All other values	

In the same way, we recode the `number of people per room \geq 6' into `1'. Click `**Add**'. Recoding for this variable is now completed.

Recode into Same Variables: Old and New Values		23
 Recode into Same Variables: Old and New Values Old Value Value: System-missing System- or user-missing Range: through Range, LOWEST through value: 	New Value Value: 1 System-missing Old> New: 0 thru 5> 0 Add Change Remove	
 Range, value through HIGHEST: 6 O All other values 		
Continue	Cancel Help	

Click **`Continue**' to return to the dialogue box at the beginning of recoding. You can choose to recode other variables. Click **`OK**' to leave after all recoding is completed.

Step 4: Principal Component Analysis (PCA)

Once the variables have been selected, we need to run a principal component analysis to create the wealth index.

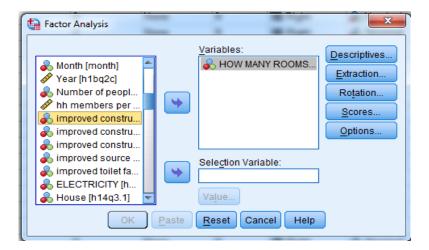
PCA is a 'data reduction' procedure. It involves replacing *many* correlated variables with a set of *principal* uncorrelated 'principal components' which can explain much of the variance and represent unobserved characteristics of the population. The objectives of a PCA are: i) to discover or reduce the dimensionality of the data set and ii) to identify new meaningful underlying variables. **The first principal component explains the largest proportion of the total variance and it is used as the wealth index to represent the household's wealth**.

In SPSS the factor analysis procedure is used to calculate the principal component. This procedure first standardizes the indicator variables by calculating the Z-scores. Then the factor coefficient scores which are also the factor loadings are generated. The indicator values are multiplied by the loadings and summed to the household wealth index. The wealth index as created is a continuous variable which can be used in correlations or regression models. The higher the score of the index, the wealthier the household.

a *uganda_fsindicators for WI_28 nov.sav [DataSet1] - IBM SPSS Statistics Data Editor												
<u>F</u> ile	<u>E</u> dit	<u>V</u> iew	<u>D</u> ata	<u>T</u> ransform	<u>A</u> nalyze	<u>G</u> raph	is <u>U</u> tilities	s Ado	d- <u>o</u> ns <u>W</u> i	ndow	<u>H</u> elp	
) 🛄			<u>p</u> orts scriptive	Statistics	* *				4
		N	ame	Label	Та	bles		•	Width	De	ecimals	Missir
1		hhid		unique hh	Co	mpare M	leans	•		2		None
2		very_h	nigh_fo	=1 if more	Ge	— neral Lin	ear Model	•		2		None
3		high_f	oodexp	=1 if more	-		d Linear Mo	dels		2		None
4		fcs_po	oor_bo			ed Mode				2		None
5		veryhi	gh_st	=1 if more	-	rrelate				2		None
6		food_o	dep	=1 if food e	-	gression		Ľ.		2		None
7		fcs_2g	groups	RECODE	-	-	1			2		None
8		WI_qu	uint	Wealth ind	-	glinear				0		None
9		exp_q	uintiles	constant_p		assify		•		0		None
10)	cross	wgt09	Weights w	_		Reduction	- Þ.	🔒 <u>F</u> acto	or		None
11	1	mult		panel weig	Sc	ale		•		2		None
12	2	pnlwg	t09	panel weig	<u>N</u> o	nparame	etric Tests	•		2		None
13	3	region		region	Fo	recas <u>ting</u>)	•		2		None
14	1	urban		urban/rural	<u>S</u> u	rvival		•		0		None
15	5	regurb)	regionxurb	Mu	Itiple Re	sponse	•		0		None
16	5	distric	t	District co	Co	mp <u>l</u> ex Sa	amples	•		0		None
17	7	h1bq2	a	Day	<u>Q</u> u	ality Con	trol	•		0		None
18	3	month	ı	Month	R	C Curve				0		None
19)	h1bq2	c	Year	Non	e	Numeric	8]	0		None
20)	hsize		Number of	p Non	e I	Numeric	9		0		None
21	1	h9q03		HOW MAN	I Non	e I	Numeric	8		0		None
22)	crowd	ina	hh momho	r Non	<u> </u>	Numoric	8		2		Nono

SPSS: <u>A</u>nalyse > <u>D</u>imension Reduction > <u>F</u>actor

The '**Factor Analysis'** dialogue box will pop up. Click to select all the variables we want to include in the factor analysis (all asset variables in this example) and press the arrow to move them into the right '**Variables**' field.



Click the **'Descriptives**' button to enter the **'Factor Analysis: Descriptives**' dialogue box. Check **Initial Solution, Coefficients, KMO and Bartlett's test of sphericity and Anti-image**. Click 'Continue' to return to the dialogue box.

Factor Analysis	Factor Analysis: Descriptives	<u> </u>
 h14q3.14: E h14q3.18: E two_poores two_poores poorest_ex poorest_Wi read_write asset_cour Percentile C two_poorest_as 	Correlation Matrix Coefficients Inverse Significance levels Reproduced Determinant Anti-image KMO and Bartlett's test of sphericity	Descriptives Extraction Rotation Scores Options

Click 'Extraction' and check Correlation Matrix, Unrotated Factor Solution, Scree Plot and Eigenvalues greater than 1. Click 'Continue'.

Factor Analysis: Extraction	1 B 100	23
Method: Principal components	~	
Analyze	Display	
Oc Correlation matrix	Unrotated factor solution	
© Covariance matrix	Scree plot	
- Extra et		
Extract		
Based on Eigenvalue Eigenvalues greater than: 1		
© Fixed number of factors		
Factors to extract:		
-		
Maximum Iterations for Converge	nce: 25	
Continue	Cancel Help	

Factor Analy	🙀 Factor Analysis: Rotation	
 h14q3.14 h14q3.18 two_pool two_pool two_pool poorest poorest 	Method O Quartimax Image: Original system Image: Original system Image: Original system Image: Or	Extraction Extraction Rotation Scores Options
Read_wrif Reset_co Percentil	Display Display Image: Related solution Image: Related solution	
& two_poor & poorest_	Maximum Iterations for Convergence: 25 Continue Cancel Help	

Click 'Rotation'. Select Varimax and Rotated Solution. Click 'Continue'.

Click 'Scores'. Select 'Save as variables' in 'Regression' method. Click 'Continue'.

 h14q3.14: D h14q3.14: D h14q3.18: D two_poorest two_poorest poorest_exp poorest_WI read_write asset_count Percentile Gr two_poorest_ass Descriptives Extraction Save as variables Method Regression Bartlett Anderson-Rubin Display factor score coefficient matrix Continue Cancel Help 	Factor Analysis		23
	 h14q3.18: D(two_poorest two_poorest poorest_exp poorest_WI read_write asset_count Percentile Gr two_poorest 	Save as variables Method Regression Bartlett Anderson-Rubin Display factor score coefficient matrix Continue Cancel	Extraction Rotation

Click 'Options'. Check 'Exclude Cases Listwise' and 'Sorted by Size'. Click 'Continue'.

Factor Analys	Factor Analysis: Options	
h14q3.14; h14q3.18; two_poor poorest_e poorest_v read_write asset_cou Percentile two_poorest_a	Missing Values © Exclude cases listwise © Exclude cases pairwise © <u>R</u> eplace with mean Coefficient Display Format Souted by size Suppress small coefficients <u>Absolute value below:</u> 10 Continue Cancel Help	Descriptives Extraction Rotation Scores Options

At this point, SPSS has completed the Principal Component Analysis.

In the SPSS output viewer⁶, we are going to encounter the correlation matrix of the selected variables. If any of the correlations are too high (above 0.9), you may need to remove one of the variables, as the two variables are likely to be measuring the same thing. Another alternative would be to combine the two variables in some way (perhaps by taking the average). If the correlations are too low, (below 0.1), you may also remove it. If others do not behave as expected, one has to investigate this and maybe remediate the problem. For example, owning a radio turned out to be lower among the wealthiest in a country A, because these households often have radio-cassette players instead. The solution was to create a new variable: "radio or radio-cassette player."

In the correlation matrix of this example, we can see that the variable 'Boat' and 'Own_cattle' may be considered for removal.

		HOW MANY										elation Matrix
		ROOMS DOES YOUR HOUSE HOLD OCCUPY?03	improved construction material of roof	improved construction material of walls	improved construction material of floor	improved source of water	improved toilet facility	ELECTRICITY	House	Other Buildings	Land	Furniture/Furn ishings
Correlation	HOW MANY ROOMS DOES YOUR HOUSE HOLD OCCUPY?03	1.000	.446	.103	.313	007	.365	.196	.179	.142	.070	.197
	improved construction material of roof	.446	1.000	.079	.382	063	.271	.197	.028	.008	.010	.266
	improved construction material of walls	.103	.079	1.000	.362	.201	006	.191	172	.147	008	.155
	improved construction material of floor	.313	.382	.362	1.000	.155	.133	.442	058	.057	089	.122
	improved source of water	007	063	.201	.155	1.000	071	.152	.008	045	024	07
	improved toilet facility	.365	.271	006	.133	071	1.000	.057	.170	.113	.065	.18
	ELECTRICITY	.196	.197	.191	.442	.152	.057	1.000	041	.004	059	.07
	House	.179	.028	172	058	.008	.170	041	1.000	032	.175	02
	Other Buildings	.142	.008	.147	.057	045	.113	.004	032	1.000	.164	.16
	Land	.070	.010	008	089	024	.065	059	.175	.164	1.000	.08
	Furniture/Furnishings	.197	.266	.155	.122	072	.185	.077	022	.167	.088	1.00
	Household Appliances e. g. Kettle, Flat iron, etc.	.214	.167	022	.217	004	.151	.339	.040	.013	049	.09
	Televsion	.206	.200	.182	.428	.113	.088	.682	.028	.036	019	.074
	Radio/ Cassette	.247	.247	.164	.243	.008	.180	.140	.033	.126	.065	.27
	Bicycle	.129	.009	.215	.083	.037	.153	027	.012	.175	.122	.23
	Motor cycle	.132	.123	.110	.178	.008	.097	.120	.064	.109	.010	.07
	Motor vehicle	.195	.112	.102	.256	.071	.055	.331	.033	.105	016	.03
	Boat	072	002	.031	064	048	083	032	.013	009	004	03
	Jewelry and Watches	.179	.154	.137	.278	.074	.035	.268	012	.074	047	.10
	Mobile phone	.288	.299	.180	.328	.029	.194	.268	.054	.133	.007	.22
	Computer	.131	.100	.077	.165	.055	.015	.305	.020	.019	.052	.02
	Other electronic equipment	.084	.098	.093	.208	.044	.019	.384	024	.023	.028	.04
	Other household assets e.g. lawn mowers, etc.	.106	.176	151	.041	166	.133	013	.066	.111	.127	.11
	own_cattle	.170	011	.113	.088	.007	.082	.053	.080	.157	006	.07

The Kaiser-Meyer-Olkin Measure of Sampling Adequacy varies between 0 and 1. The values that are closer to 1 are better. A value of **0.6** is a suggested minimum acceptable value. In our example, we have a value of 0.803, which is satisfactory.

⁶ See more interpretation of SPSS PCA outputs at http://statistics.ats.ucla.edu/stat/spss/output/principal components.htm.

Creation of a Wealth Index - 19 -

KMO and Bartlett's Test

Kaiser-Meyer-Olkin	Kaiser-Meyer-Olkin Measure of Sampling				
Adequacy.	<mark>.803</mark>				
	Approx. Chi-Square	14155065.78 0			
Bartlett's Test of					
Sphericity	df	276			
	Sig.	.000			

The Component 1 is used as the wealth index as it accounts for the largest proportion of the variance. In the example dataset, the PCA generates a variable labelled as 'REGR factor score 1 for analysis 1' which is the wealth index.

		Initial Eigenvalu	ies	Extraction	n Sums of Square	ed Loadings
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.623	30 1 8 9	30.189	3.623	30.189	30.189
2	1.464	12198	42.387	1.464	12.198	42.387
3	1.025	8.546	50.932	1.025	8.546	50.932
4	.938	7.814	58.746			
5	.887	7.395	66.141			
6	.840	7.001	73.142			
7	.725	6.045	79.188			
8	.679	5.656	84.844			
9	.593	4.945	89.789			
10	.545	4.545	94.334			
11	.387	3.224	97.559			
12	.293	2.441	100.000			

Total Variance Explained

Extraction Method: Principal Component Analysis.

Initial Eigenvalues: 'Eigenvalues are the variances of the principal components. Because we conducted our principal components analysis on the correlation matrix, the variables are standardized, which means that each variable has a variance of 1, and the total variance is equal to the number of variables used in the analysis, in this case, 12'.

Total: 'This column contains the Eigenvalues. The first component will always account for the most variance (and hence have the highest Eigenvalue)'.

Extraction sums of squared loadings: 'The three columns of this half of the table exactly reproduce the values given on the same row on the left side of the table. The number of rows reproduced on the right side of the table is determined by the number of principal components whose Eigenvalues are 1 or greater'. (Sources: IDRE, UCLA)

In the output, we can also see the component matrix. This table contains the component loadings, which are the correlations between the variable and the component. Possible values range from -1 to +1.

Component Matrix ^a								
	Component							
	1 2 3							
HOW MANY ROOMS DOES YOUR HOUSE HOLD OCCUPY?03	.353	.553	271					
improved construction material of roof	.543	.370	217					
improved construction material of walls	.473	090	.618					
improved construction material of floor	.757	151	.186					
ELECTRICITY	.732	438	146					
Furniture/Furnishings	.298	.524	104					
Household Appliances e. g. Kettle, Flat iron, etc.	.570	275	451					
Televsion	.722	404	115					
Radio/ Cassette	.442	.480	.108					
Motor cycle	.302	.132	.464					
Jewelry and Watches	.484	004	.001					
Mobile phone								
Extraction Method: Principal a. 3 components extracte		Analysis.						

****** The construction of the wealth index is an **iterative process**. To obtain the best results, we usually need to conduct a few rounds of PCA including or excluding certain variables based on the factor coefficient scores we see in the PCA outputs.

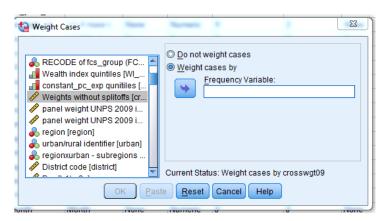
Step 5: Create wealth index quintiles

Before we rank the wealth index and create the quintiles, we need to turn weights on to correct over/under sampling of any group.

<u>F</u> ile	<u>E</u> dit	View	<u>D</u> ata	Transform	<u>A</u> nalyze	Graphs	<u>U</u> tilities	Add- <u>o</u> ns	Windo	ow <u>H</u> elp			
E			7	Define <u>V</u> ariable	Propertie	S		M			- 4 <u>2</u> -		0
_			34	Set Measureme	ent <u>L</u> evel fo	or Unknow	/n						
1		hhid N		<u>C</u> opy Data Prop	perties			Width 10		Decimals	Missing None	Columns 20	
2		very_t	1	New Custom A	ttri <u>b</u> ute			10		2	None	8	IIII Ri
3		high 1	8	Define Dates				9 9		2	None	8	三 Ri
4		fcs p		– Define Multiple	Response	e Sets		9		2	None	8	E Ri
		veryhi		dentify Duplica				9		2	None	8	I Ri
6		food			10 04363			9		2	None	8	遍 Ri
7		fcs 2	_	Sort Cases				19		2	None	8	≣ Ri
8	}	WI_q		Sort Varia <u>b</u> les				3		0	None	5	圖 Ri
9)	exp_c		Tra <u>n</u> spose				3	(0	None	10	≣ Ri
1	0	cross	1	Mer <u>g</u> e Files			•	Э	:	2	None	8	🗃 Ri
1	1	mult	₩.	Restructure				Ð	1	2	None	8	🗏 Ri
1:	2	pnlwg	a	Aggregate				Э	:	2	None	8	彊 Ri
1:	3	regior 📆 Copy Dataset urban 📰 Split File regurt 🗰 Select Cases						8		2	None	8	🗃 Ri
14	4							В	(0	None	8	🗏 Ri
1	5							В		0	None	8	圖 Ri
10	6	distric district Cases					10	(0	None	8	🗐 Ri	
1	7	h1bq2	- I.	eight Cases				8	(0	None	8	🗐 Ri
1	8	month		Month	None	Nu	imeric	8	(0	None	8	圖 Ri
19	9	h1bq2	с	Year	None	Nu	meric	8		0	None	8	疆 Ri
2	-	hsize		Number of			meric	9		0	None	8	遍 Ri
2		h9q03		HOW MAN			meric	8		0	None	8	🗏 Ri
2:	2	crowd	na	hh membe	r None	Nu	meric	8		2	None	10	彊 Ri
Data	View	Variab											

SPSS: Data→ Weight Cases

In the dialogue box that pops up, select the option **`Weight cases by'**. Click on the weighting variable in the left field and click on the right arrow to move this variable into the right field called **`Frequency Variable'**.



In order to better understand the wealth index, which is a continuous variable, it is useful to recode the index into a categorical variable. The best way to do it is to rank the WI (the first variable created from the PCA) into deciles or quintiles, dividing all households into five or ten equal groups. In the SPSS demonstration, we rank the WI into quintiles.

SPSS: Transform > Rank Cases

ile <u>E</u> o	dit <u>V</u>	<u>′</u> iew <u>D</u> ata	<u>T</u> ransform	Analyze	<u>G</u> raphs	<u>U</u> tilities	Add- <u>o</u> r	is <u>W</u> indow	<u>H</u> elp					
			📑 <u>C</u> ompu 🚮 Count V							st 📰		ABS		
		Name			in Gases.				Values	Туре	Width	Decimals	Missing	
1		nhid	Shi <u>f</u> t Va						None	Numeric	10	2	None	2
2	١	/ery_high_fo.	🔤 Recode	e into <u>S</u> ame	e Variable:	S	1	to food	None	Numeric	9	2	None	8
3	ł	nigh_foodexp	🔤 <u>R</u> ecode	e into Differ	ent Variab	les		to food	None	Numeric	9	2	None	8
4	f	cs_poor_bo.	🛐 <u>A</u> utoma	itic Recode					None	Numeric	9	2	None	8
5	١	/eryhigh_st	📴 Visual 🖪	Binning					None	Numeric	9	2	None	8
6	f	ood_dep	Rank C	ases					None	Numeric	9	2	None	8
7	f	cs_2groups	Date an		zord			erline and ad	{1.00, P	Numeric	19	2	None	8
8	١	WI_quint							None	Numeric	3	0	None	5
9		exp_quintiles	Create	_					None	Numeric	3	0	None	1
10	(crosswgt09	📲 Replac	e Missing <u>)</u>	/alues				None	Numeric	9	2	None	8
11	-	mult	🍘 Randor	m Number	<u>G</u> enerato	rs			None	Numeric	9	2	None	8
12		onlwgt09	Run Pe	nding <u>T</u> rar	sforms	Ctrl+(G		None	Numeric	9	2	None	8
13	1	region	region						{.00, ka	Numeric	8	2	None	8
14		urban	urban/rural identifier					{0, rural}	Numeric	8	0	None	8	
15	1	regurb	regionxurban - subregions					{1, kamp	Numeric	8	0	None	8	
16	(district	District code					None	Numeric	10	0	None	8	
17	ł	n1bq2a	Day					None	Numeric	8	0	None	8	
18		month	Month	Month					None	Numeric	8	0	None	8
19	I	n1bq2c	Year						None	Numeric	8	0	None	8
20		nsize	Number of	people in	househol	d			None	Numeric	9	0	None	8
21		19q03	HOW MANY ROOMS DOES YOUR HOUSE HOLD OCCUP					None	Numeric	8	0	None	8	
22		crowding	hh members per room					None	Numeric	8	2	None	1	
23	i	mproved_roof	improved construction material of roof				{.00, no}	Numeric	8	2	None	1		
24	i	mproved_w	improved construction material of walls					{.00, no}	Numeric	8	2	None	1	
25	i	mproved_floor	improved construction material of floor					{.00, no}	Numeric	8	2	None	1	
26	i	mproved_w	improved s	ource of w	ater				{.00, no}	Numeric	8	2	None	1
		•									-	-		-
Data Vie	ew v	ariable View												

Select `wealth index' in the left field and click the arrow to move it to the right `**Variable(s)**' field. Then click `**Rank Types**' to open the dialogue box. Check `**Rank**' and `**Ntiles**'. Set the number of Ntiles at **5** as we are ranking in **quintiles**.

🙀 Rank Cases	pe	Numeric	э	2	ivone
	he	Numeric	9	2	None
Montorized_vehicle Variable(s): Variable(s): Anh Types		Rank Cases: Ty	pes		×
δ h14q3.8: Does a Τies δ h14q3.9: Does a		Rank		ractional rank a	_
Antiq3.14: Does By: Antiq3.18: Does By:		<u>Savage scor</u> <u>F</u> ractional ra		Sum of <u>c</u> ase wei <u>I</u> tiles: 5	gnts e
two_poorest_exp	1	Proportion e	stimates 🥅 N	lormal scores	9
Assign Rank 1 to	Ì ⊢F	Proportion Est	imation Formu	la	2
Smallest value	0	● <u>B</u> lom © <u>T</u> i	ukey © Ran <u>k</u>	it loo <u>V</u> an der Wa	aerden
© Largest value	r	Conti	nue Cance	Help	2
OK Paste Reset Cancel Help	ne	Numeric	10	0	None
h1bq2a Day	None	Numeric	8	0	None
month Month	None	Numeric	8	n	None

Click 'Continue' and 'OK'. Two variables will be created and appear in the dataset. The variable called "percentile group of WI" is our wealth quintile variable.

Created variables ^a								
Source	Function	New variable	Label					
variable								
t	Rank	RAN001	Rank of WI					
WI ^b	Percentile group ^c	NTI001	Percentile group of WI					

a. Mean rank of tied values is used for ties.

b. Ranks are in ascending order.

c. 5 groups are generated.

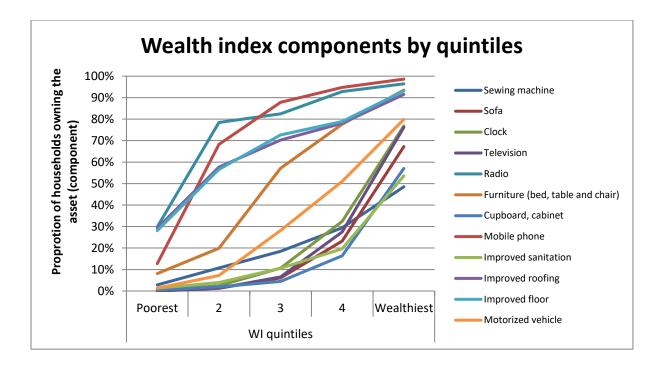
Step 6. Graph the wealth index

Graphing these quintiles or deciles by the variables included in the PCA will help the analyst to have an immediate idea about whether those variables are appropriate for the wealth index construction. Analysts can perform this step after running the PCA as a final check if the most desirable variables have been included in the wealth index.

To create the graph, run a cross-tabulation between the new categorical wealth index quintiles or deciles and the variables (assets and housing characteristics used in the analysis). This will show the prevalence of households that own the selected assets in each quintile. There should be a positive relationship between the independent and dependent variable. If there are variables included where this is not the case, they should be examined further and considered for removal.

In addition, the analysts may find it very useful to do this cross-tabulation by urban and rural areas separately as a double check on how this national index applies to both places of residence. If the variables in either rural or urban areas show insignificant or opposite patterns from what we expect, we may consider reconstructing the index or creating separate ones for both of them.

In the example graph below, ownership of every variable included in the PCA increases as the quintiles go from poorest to wealthiest. This indicates that the variables included for the PCA are appropriate.



Step 7. Select the final result and report the variables

After PCA procedures and wealth index graphing, we will have a good idea of the most appropriate wealth index. A rule of thumb to understand if the index created is appropriate is to run a correlation between the two latest first factors (of 2 different principal component analyses). If their correlation coefficient is close to 1 (0.998/0.999) then the two indicators are very similar and the wealth indices are also very similar.

When a satisfactory wealth index has been created this can be used for further analysis in relation to other indicators.

Additional Reading

Deaton and Zaidi, 2001: "Guidelines for Constructing Consumption Aggregates

For Welfare Analysis" Washington, DC: World Bank (2002)

Filmer and Prichett: "Estimating Wealth Effects without Expenditure Data -- or Tears:

An Application to Educational Enrollments in States of India" Demography. Feb;38(1):115-32.(2001)

Fimler and Scott, 2008: "Assessing Asset Indices" Policy Research Working Paper 4605, WB

Howe, L.D., Heargreave, J.R., Gabrysch S., and Huttly. S.R.A., 2009: "Is the Weath Index a Proxy for Consumption Expenditure? A Systematic Review" Journal of Epidemiol Community Health, 2009: 63, 871-880

Rutstein, 2008 : «The DHS Wealth Index: Approaches for Rural and Urban Areas" DHS Working Papers No. 60

McKenzie, D.: "Measuring Inequality with Asset Indicators" Journal of Population Economics 18(2): 229-260 (2005).