



Práce s daty SHARE v programu SPSS

Workshop, 19.4.2017

PhDr. Jaroslava Pospíšilová

Obsah přednášky

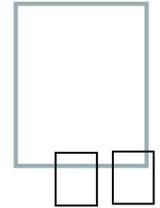


1. Obecné seznámení s výzkumem SHARE
2. Data *easySHARE*
3. Jednoduché analýzy dat *easySHARE* v SPSS

Výzkum SHARE

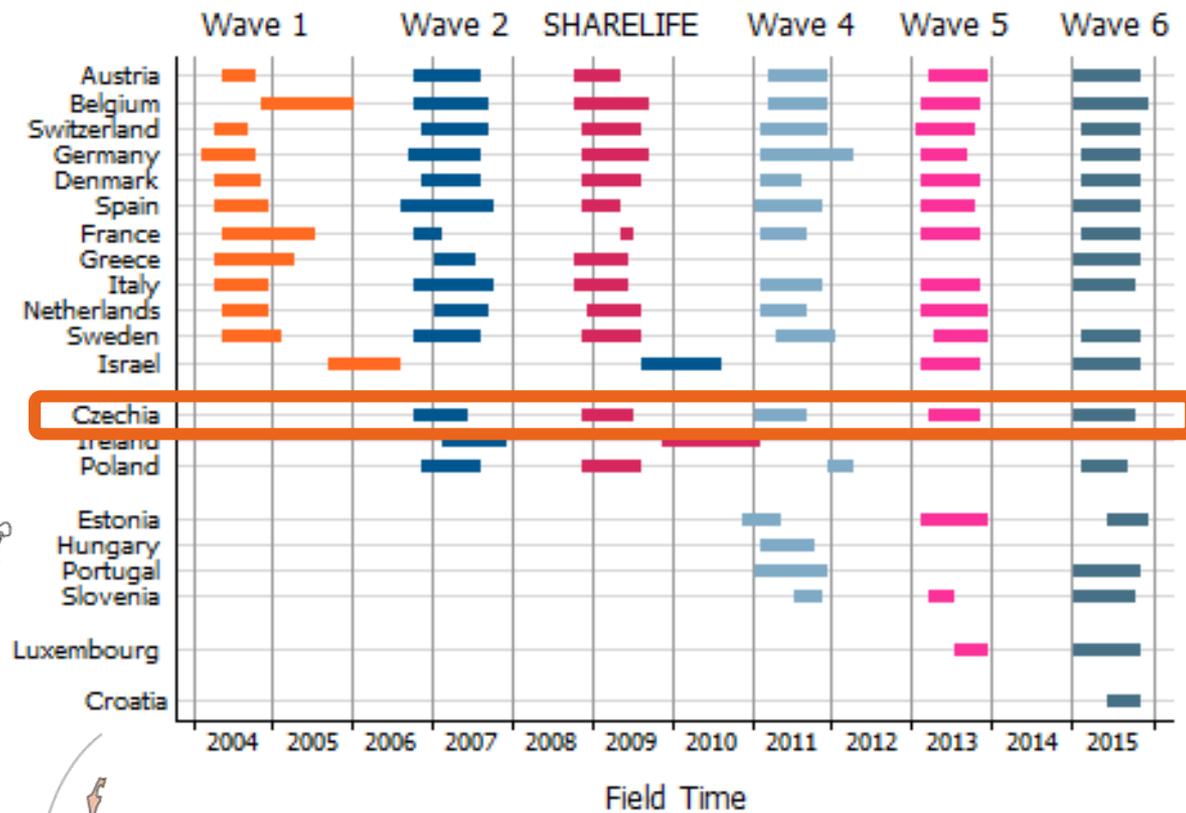
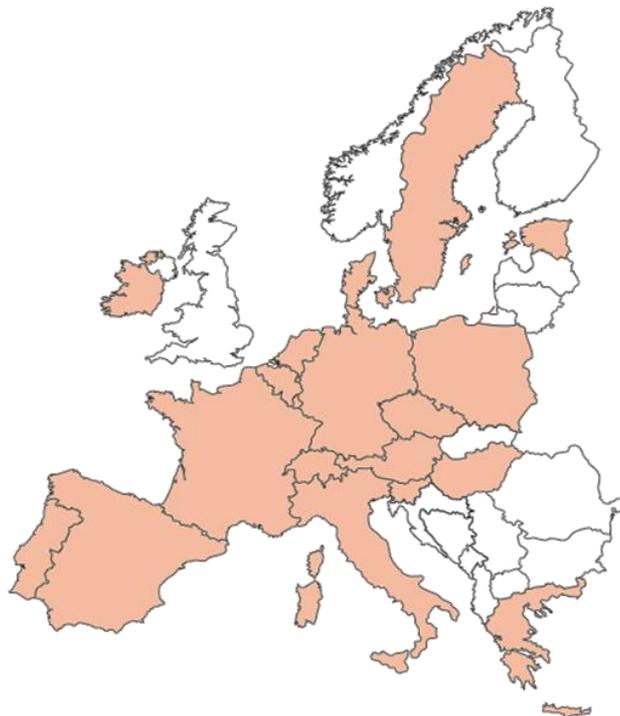
▶ **Vhodný respondenti:**

- osoby 50+ let a jejich partneři (ti bez ohledu na věk)



- ▶ **Panelový výzkum**, snaha o co nejvyšší návratnost
- ▶ Vzorek pravidelně doplňován metodou náhodného pravděpodobnostního výběru domácností
- ▶ **Metoda sběru:** CAPI, případně pro DROP-OFF dotazník metoda PAPI
- ▶ V ČR sběr zajišťuje agentura SC&C

Účast ve vlnách



Speciální data

Special Data Sets

Apart from the regular panel and the retrospective life histories (SHARELIFE), SHARE offers a variety of additional datasets:

1. *easy*SHARE: Our training dataset designed for teaching purposes at universities and for beginners to get a first glance into SHARE.

[>> more](#)

2. Record Linkage: Country specific projects to link SHARE data with selected administrative records.

[>> more](#)

3. Interviewer Survey: A special dataset based on a questionnaire targeted at SHARE's interviewers.

[>> more](#)

4. Job Episodes Panel: A retrospective panel about working life histories of SHARELIFE respondents.

[>> more](#)

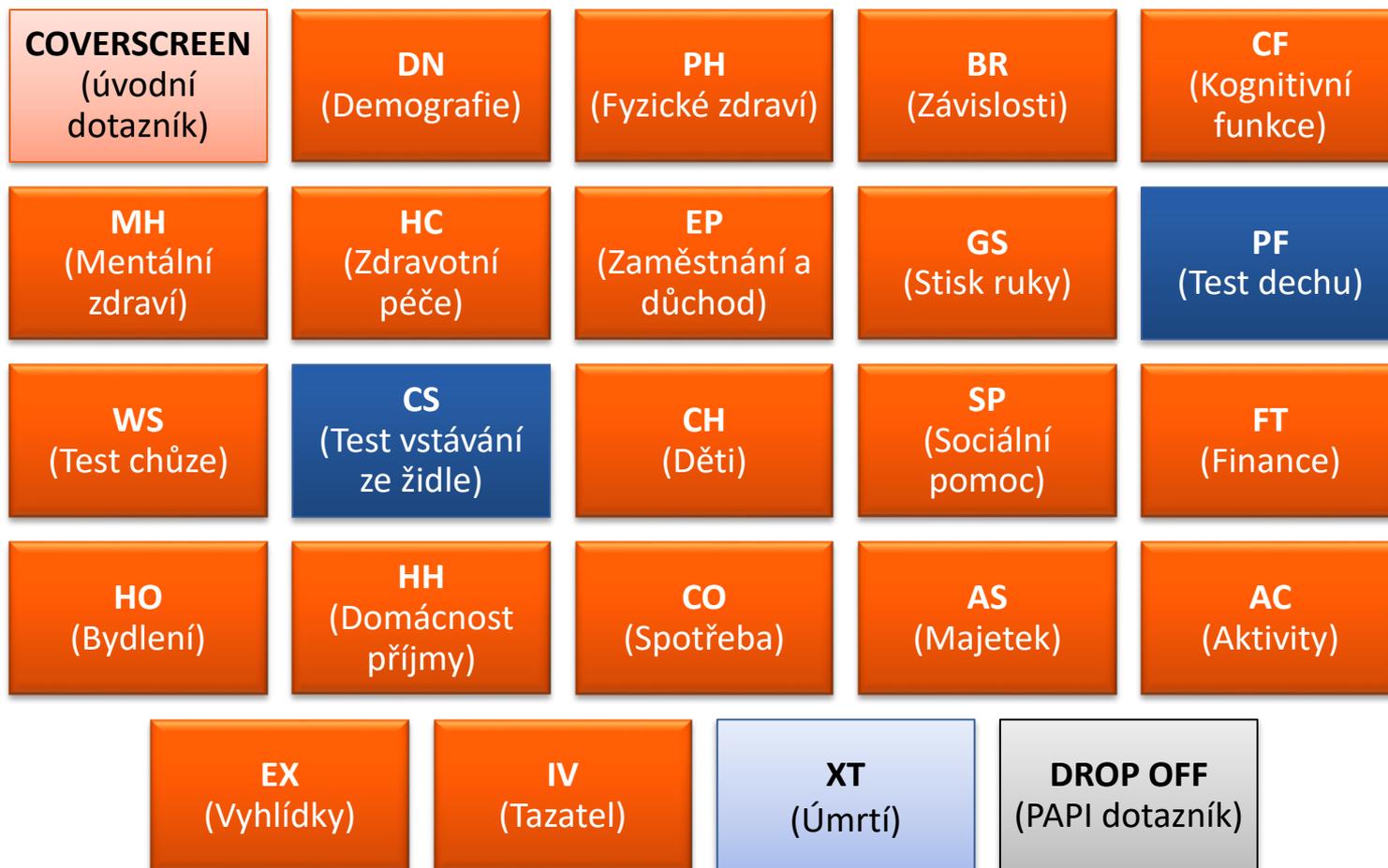
5. Biomarkers: Objective health measures included in SHARE.

[>> more](#)

Moduly 1. vlny

COVERSCREEN (úvodní dotazník)	DN (Demografie)	PH (Fyzické zdraví)	BR (Závislosti)	CF (Kognitivní funkce)
MH (Mentální zdraví)	HC (Zdravotní péče)	EP (Zaměstnání a důchod)	GS (Stisk ruky)	WS (Test chůze)
CH (Děti)	SP (Sociální pomoc)	FT (Finance)	HO (Bydlení)	HH (Domácnost příjmy)
CO (Spotřeba)	AS (Majetek)	AC (Aktivity)	EX (Vyhlídky)	IV (Tazatel)

Moduly 2. vlny



Moduly 3. vlny - SHARELIFE

Questionnaire Module	Variable name prefix	Content	Variables deviating from module name
ac	sl_ac	Accommodation section	
cs	sl_cs	Childhood section	
dq	sl_dq	Disability	
fs	sl_fs	Financial history section	sl_hh017_
gl	sl_gl	General life questions	
gs	sl_gs	Grip strengths	
hc	sl_hc	Childhood health care	
hs	sl_hs	Childhood health section	sl_ph003_
iv	sl_iv	Interviewer	
rc	sl_rc	Retrospective children	
re	sl_re	Work history	
rp	sl_rp	Partner section	
st	sl_st	Demographics	
wq	sl_wq	Work quality	
xt	sl_xt	End of life interview	

Moduly 4. vlny

COVERSCREEN (úvodní dotazník)	DN (Demografie)	SN (Sociální sítě)	CH (Děti)	PH (Fyzické zdraví)
BR (Závislosti)	CF (Kognitivní funkce)	MH (Mentální zdraví)	HC (Zdravotní péče)	EP (Zaměstnání a důchod)
GS (Stisk ruky)	PF (Test dechu)	SP (Sociální pomoc)	FT (Finance)	HO (Bydlení)
HH (Domácnost příjmy)	CO (Spotřeba)	AS (Majetek)	AC (Aktivity)	EX (Vyhledky)
	IV (Tazatel)	XT (Úmrtí)	DROP OFF (PAPI dotazník)	

Moduly 5. vlny

COVERSCREEN (úvodní dotazník)	DN (Demografie)	SN (Sociální sítě)	CH (Děti)	PH (Fyzické zdraví)
BR (Závislosti)	CF (Kognitivní funkce)	MH (Mentální zdraví)	HC (Zdravotní péče)	EP (Zaměstnání a důchod)
IT (Práce na počítači)	MC (Vlastní dětství)	GS (Stisk ruky)	CS (Test vstávání)	SP (Sociální pomoc)
FT (Finance)	HO (Bydlení)	HH (Domácnost)	CO (Spotřeba)	AS (Majetek)
AC (Aktivity)	EX (Vyhlídky)	IV (Tazatel)	XT (úmrtí)	DROP OFF (PAPI dotazník)

Moduly 6. vlny

COVERSCREEN (úvodní dotazník)	DN (Demografie)	SN (Sociální sítě)	CH (Děti)	PH (Fyzické zdraví)
BR (Závislosti)	CF (Kognitivní funkce)	MH (Mentální zdraví)	HC (Zdravotní péče)	EP (Zaměstnání a důchod)
IT (Práce na počítači)	BS (Vzorky krve)	GS (Stisk ruky)	PF (Test dechu)	SP (Sociální pomoc)
FT (Finance)	HO (Bydlení)	HH (Domácnost)	CO (Spotřeba)	AS (Majetek)
AC (Aktivity)	EX (Vyhlídky)	IV (Tazatel)	XT (úmrtí)	DROP OFF (PAPI dotazník)

Porovnání proměnných napříč vlnami

The screenshot shows the SHARE project website. The main header features the SHARE logo (Survey of Health, Ageing and Retirement in Europe 50+ in Europe) and a map of Europe. The page title is 'Cross Wave Comparison'. The left navigation menu includes 'Data Access & Documentation', 'Documentation', and 'Cross Wave Comparison'. The main content area explains the purpose of the comparison and provides a table of explanatory notes. A red arrow points from the 'Cross Wave Comparison' menu item to the 'Download' link.

Select Country

SHARE
Survey of Health, Ageing and Retirement in Europe
50+ in Europe

Survey of Health, Ageing and Retirement in Europe

You are here: Data Access & Documentation - Documentation - Cross Wave Comparison

Cross Wave Comparison

For the correspondence across waves the basic rule holds that questions are identically named if the content has not changed.

As changes sometimes have to be conducted due to societal changes (eg. new communication tools) or due to flaws found in the previous waves (e.g. ambiguous understanding of question wording) you will find an overview of the generic versions of the questionnaires in wave 1, wave 2, wave 4, and wave 5 where you can track the deviations between waves easily. "Wave 3" was a different questionnaire named SHARELIFE and is not included.

This overview contains:

- the comparison of question texts, response options and interviewer instructions in the main questionnaires
- the comparison of drop-off questionnaires. We also documented movements from drop-off to the CAPI questionnaire.

Download: [Cross wave comparison-document](#).

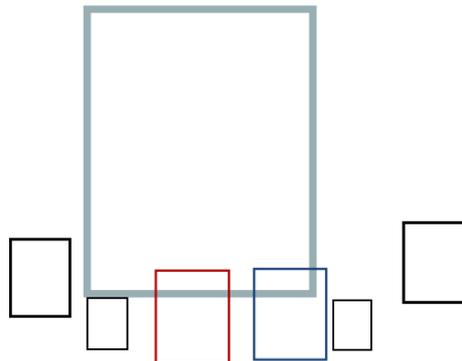
Explanatory notes:

W#_Question_Text	Question texts of wave 1, 2, 4 or 5
W#_Response_Options	Response options of wave 1, 2, 4 or 5
W#_Interviewer_Instruction	Interviewer instructions of wave 1, 2, 4 or 5
.	Question has not been asked at all in corresponding wave
[none]	Question has been asked in corresponding wave but no interviewer instruction, response option or in some cases question text (e.g. interviewer observation) was given

Please be aware that all tables are still under construction and will be updated permanently.

For feedback please contact us at info@share-project.org.

Vlna 5 - moduly



Overview: Who Answers What in the CAPI Questionnaire?

CAPI Module	Name	All respondents ¹	Household Resp.	Financial Resp.	Family Resp.	non-proxy
CV_R	Coverscreen	cv-respondent				
DN	Demographics	X				
CH	Children				x	
PH	Physical Health	x				
BR	Behavioural Risks	x				
CF	Cognitive Function	x				x
MH	Mental Health	x				x (partly)
HC	Health Care	x				
EP	Employment and Pensions	x				
IT	IT module	x				
MC	Mini Childhood	x				
GS	Grip Strength	x				x
CS	Chair Stand	x				x
SP	Social Support	x (partly)			x (partly)	
FT	Financial Transfers			x		
HO	Housing		x			
HH	Household Income		x			
CO	Consumption		x			
AS	Assets			x		
AC	Activities	x				x
EX	Expectations	x				x
IV	Interviewer Observations	interviewer				
XT	End-of-Life Interview	proxy respondents				

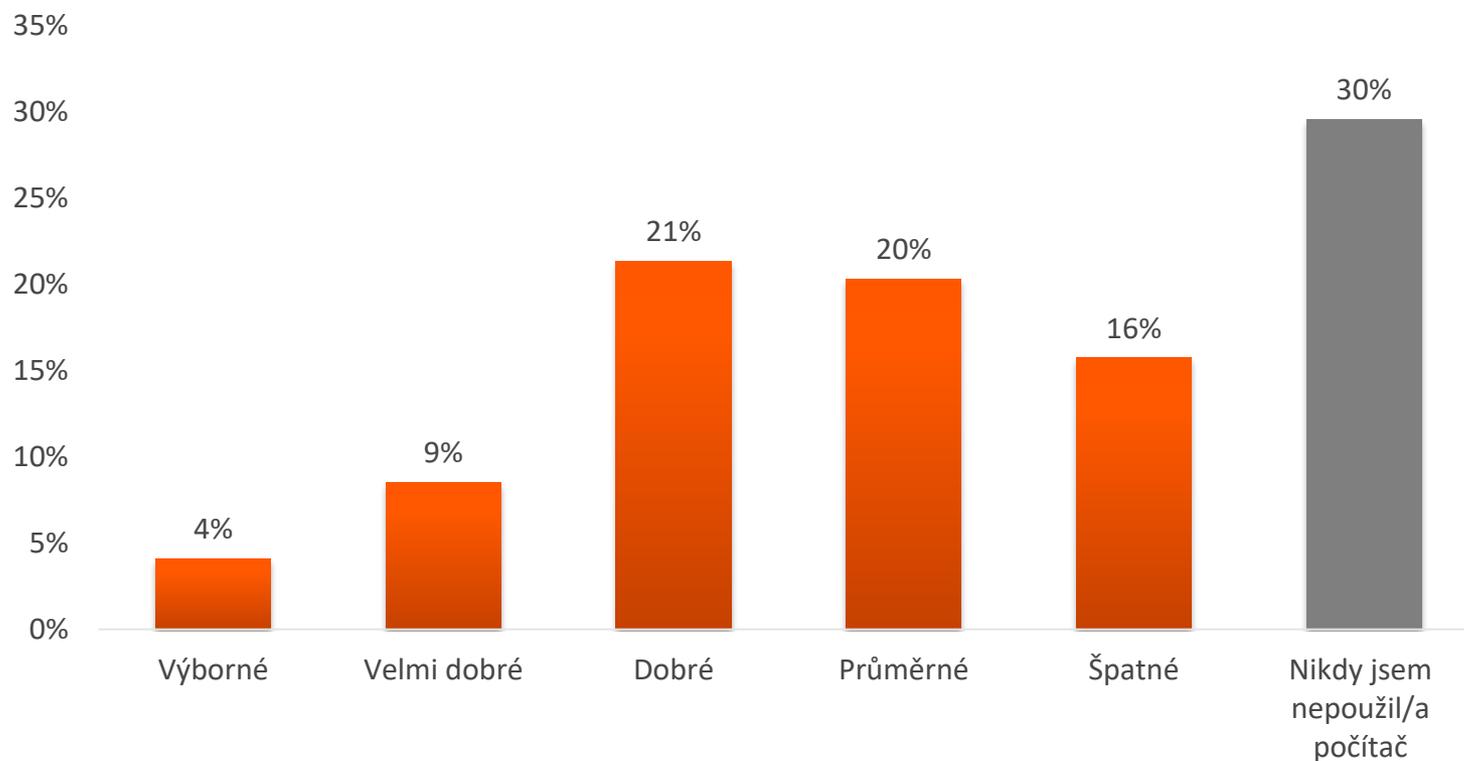
¹ "All respondents" only refers to types of respondents here. Due to other filters/routing not all modules are answered by all respondents.

Praktický příklad č. 1 – IT schopnosti

- ▶ **it003** *_ Jak byste ohodnotil/a Vaše počítačové dovednosti? Řek l/a byste, že jsou...*
 - 1. Výborné
 - 2. Velmi dobré
 - 3. Dobré
 - 4. Průměrné
 - 5. Špatné
 - 6. Nikdy jsem nepoužil/a počítač (POUZE SPONTÁNNÍ ODPOVĚDI)

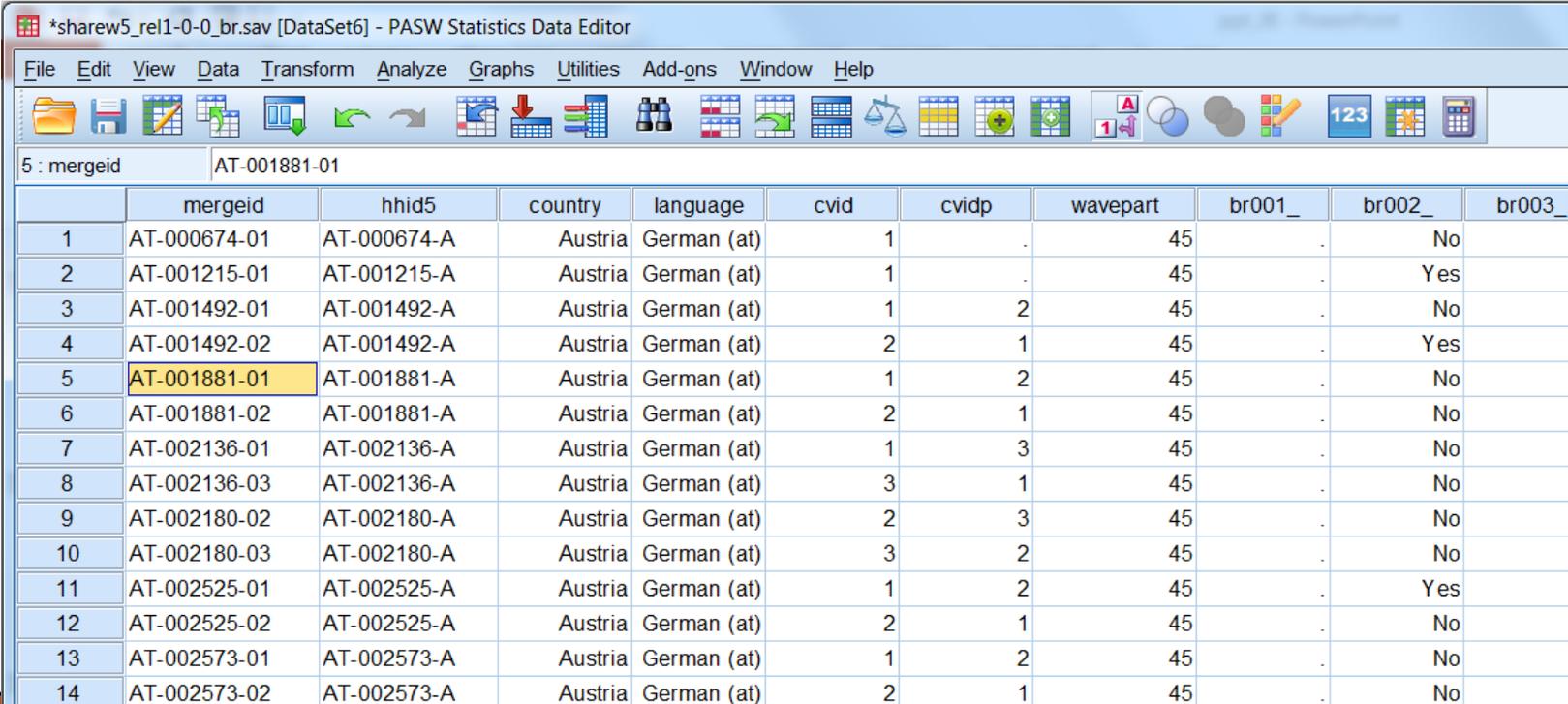
Praktický příklad č. 1 – řešení

Počítačové schopnosti Evropanů 50+



Pozn. identifikátory

- ▶ **mergeid** – unikátní kód respondenta (napříč vlnami)
- ▶ **hhid** – unikátní kód domácnosti (napříč vlnami)



*sharew5_rel1-0-0_br.sav [DataSet6] - PASW Statistics Data Editor

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

5 : mergeid AT-001881-01

	mergeid	hhid5	country	language	cvid	cvidp	wavepart	br001_	br002_	br003_
1	AT-000674-01	AT-000674-A	Austria	German (at)	1	.	45	.	No	
2	AT-001215-01	AT-001215-A	Austria	German (at)	1	.	45	.	Yes	
3	AT-001492-01	AT-001492-A	Austria	German (at)	1	2	45	.	No	
4	AT-001492-02	AT-001492-A	Austria	German (at)	2	1	45	.	Yes	
5	AT-001881-01	AT-001881-A	Austria	German (at)	1	2	45	.	No	
6	AT-001881-02	AT-001881-A	Austria	German (at)	2	1	45	.	No	
7	AT-002136-01	AT-002136-A	Austria	German (at)	1	3	45	.	No	
8	AT-002136-03	AT-002136-A	Austria	German (at)	3	1	45	.	No	
9	AT-002180-02	AT-002180-A	Austria	German (at)	2	3	45	.	No	
10	AT-002180-03	AT-002180-A	Austria	German (at)	3	2	45	.	No	
11	AT-002525-01	AT-002525-A	Austria	German (at)	1	2	45	.	Yes	
12	AT-002525-02	AT-002525-A	Austria	German (at)	2	1	45	.	No	
13	AT-002573-01	AT-002573-A	Austria	German (at)	1	2	45	.	No	
14	AT-002573-02	AT-002573-A	Austria	German (at)	2	1	45	.	No	

Otázky před analýzou



Přístup k datům

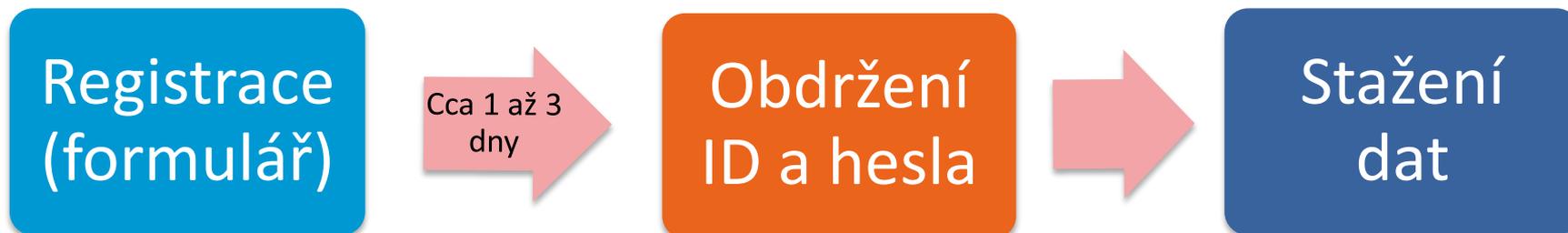
Co řešit ještě než začneme s analýzou

Co je *easySHARE*

Dostupnost SHARE dat

- ▶ Pro nekomerční účely data SHARE **volně k dispozici**
- ▶ V současnosti k dispozici¹:
 - data z **6 vln**
 - **easySHARE** (zjednodušená data ze 4 vln obsahující výběr základních proměnných)
- ▶ Dostupné ve formátech pro programy
 - STATA
 - SPSS
 - částečně R

Postup při stažení dat SHARE



Registrace

- ▶ Ze stránek <http://www.share-project.org> si nejprve stáhněte registrační formulář



The screenshot shows a web browser window with the URL www.share-project.org/data-access-documentation/research-data-center-data-access.html. The page features the SHARE logo and a navigation menu on the left. The main content area is titled "Research Data Center & Data Access" and includes a search bar and a list of conditions for data access. A blue callout bubble labeled "Formulář" points to the text "The required forms can be downloaded here." in the second condition. An orange callout bubble labeled "Cesta" points to the "Data Access & Documentation" link in the left sidebar.

Select Country

SHARE
Survey of Health, Ageing
and Retirement in Europe
50+ in Europe

Survey of Health,
Ageing and
Retirement in
Europe

You are here: Data Access & Documentation · Research Data Center & Data Access

Research Data Center & Data Access

The SHARE data are distributed through our Research Data Center which is physically located at CentERdata on the Tilburg University campus in the Netherlands. The SHARE Research Data Center complies with the *Criteria of the German Council for Social and Economic Data* for providing access to microdata.

SHARE Data Access Rules

The SHARE data can be **downloaded** from the SHARE Research Data Center under the following conditions:

1. Applicants must have a scientific affiliation and have to sign a statement confirming that under no circumstances the data will be used for other than purely scientific purposes.
2. Data will only be made available after these documents have been received, by mail or fax (care of Josette Janssen; address: CentERdata, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands; e-mail: jjanssen@uvt.nl; fax: +31 13 4662764). **The required forms can be downloaded here.** Upon request a copy by fax can be sent. Upon receipt of the signed statements users will receive a username and password enabling them to download the data. Registration is free of charge. Registered users are allowed to use data of the SHARE project as long as the scientific affiliation indicated in the user statement is valid. The original login code and password persist for all subsequent releases of the data. A new statement has to be filled, however, when **any** of the specifications given in the statement (incl. e-mail address) change. If users forgot their password, they should go to <http://link.centerdata.nl/sharedata>. A password can however only be reset if the e-mail address typed in is the same as used when applying for the data. Additionally, the data are available via the *GESIS Data Archive for the Social Sciences*.

Home

Data Access & Documentation

Research Data Center & Data Access

easySHARE

Record Linkage (SHARE-RV)

Documentation

Questionnaires

Sample

User Statistics

Methodological Research

Press Information & News

Contact & Organisation

Publications

Search

Formulář

Cesta

Formulář

- ▶ Vyplněný a naskenovaný formulář odeslat na jjanssen@uvt.nl
- ▶ Za 1 až 3 dny přijde ID a přístupové heslo



Statement concerning the use of SHARE data

The undersigned (*please use capital letters*):

(last name, first name/s)

(position)

(scientific affiliation)

hereby undertakes to carry out work on data of the SHARE project in accordance with the following conditions only.

- (1) He/she undertakes **neither** to make copies of the data available to others **nor** to enable any third party access to the database.
- (2) He/she undertakes to use the data for scientific research **only**. Commercial use of the SHARE data is strictly prohibited.
- (3) He/she undertakes to take **no** action aiming at a re-identification of participants.
- (4) He/she undertakes only to use data of the SHARE project **as long as** the affiliation indicated is valid. Scientific affiliation is required. A new statement has to be filled when **any** of the specifications (incl. e-mail address) given in this statement change.
- (5) The undersigned undertakes to **provide a reference** of **all** publications based on the data to the SHARE co-ordination team (e-mail: info@share-project.org) and to include the following **acknowledgement** in all publications:

"This paper uses data from SHARE wave 5 release 1.0.0, as of March 31st 2015 (DOI: 10.6103/SHARE.w5.100) or SHARE wave 4 release 1.1.1, as of March 28th 2013 (DOI: 10.6103/SHARE.w4.111) or SHARE waves 1 and 2 release 2.6.0, as of November 29th 2013 (DOIs: 10.6103/SHARE.w1.260 and 10.6103/SHARE.w2.260) or SHARELIFE release 1.0.0, as of November 24th 2010 (DOI: 10.6103/SHARE.w3.100). The SHARE data collection has been primarily funded by the European Commission through the 5th Framework Programme (project QLK6-CT-2001-00360 in the thematic programme Quality of Life), through the 6th Framework Programme (projects SHARE-I3, RII-CT-2006-062193, COMPARE, CIT5-CT-2005-028857, and SHARELIFE, CIT4-CT-2006-028812) and through the 7th Framework Programme (SHARE-PREP, N° 211909, SHARE-LEAP, N° 227822 and SHARE M4, N° 261982). Additional funding from the U.S. National Institute on Aging (U01 AG09740-13S2, P01 AG005842, P01 AG08291, P30 AG12815, R21 AG025169, Y1-AG-4553-01, IAG BSR06-11 and OGHA 04-064) and the German Ministry of Education and Research as well as from various national sources is gratefully acknowledged (see www.share-project.org for a full list of funding institutions)."

The acknowledgement included has to refer to the release the publication is based on (see www.share-project.org for an up-to-date version of the acknowledgement when new releases are available).

- (6) He/she agrees to be included in the list of users of the SHARE project and to be informed about **updates of the data** via e-mail.
- (7) He/she agrees to receive the SHARE **newsletter** via e-mail. The newsletter is published twice a year and contains information on the SHARE project. Unsubscribing is possible at any time by sending an e-mail to info@share-project.org (subject: "Unsubscribe SHARE newsletter").

(signature)

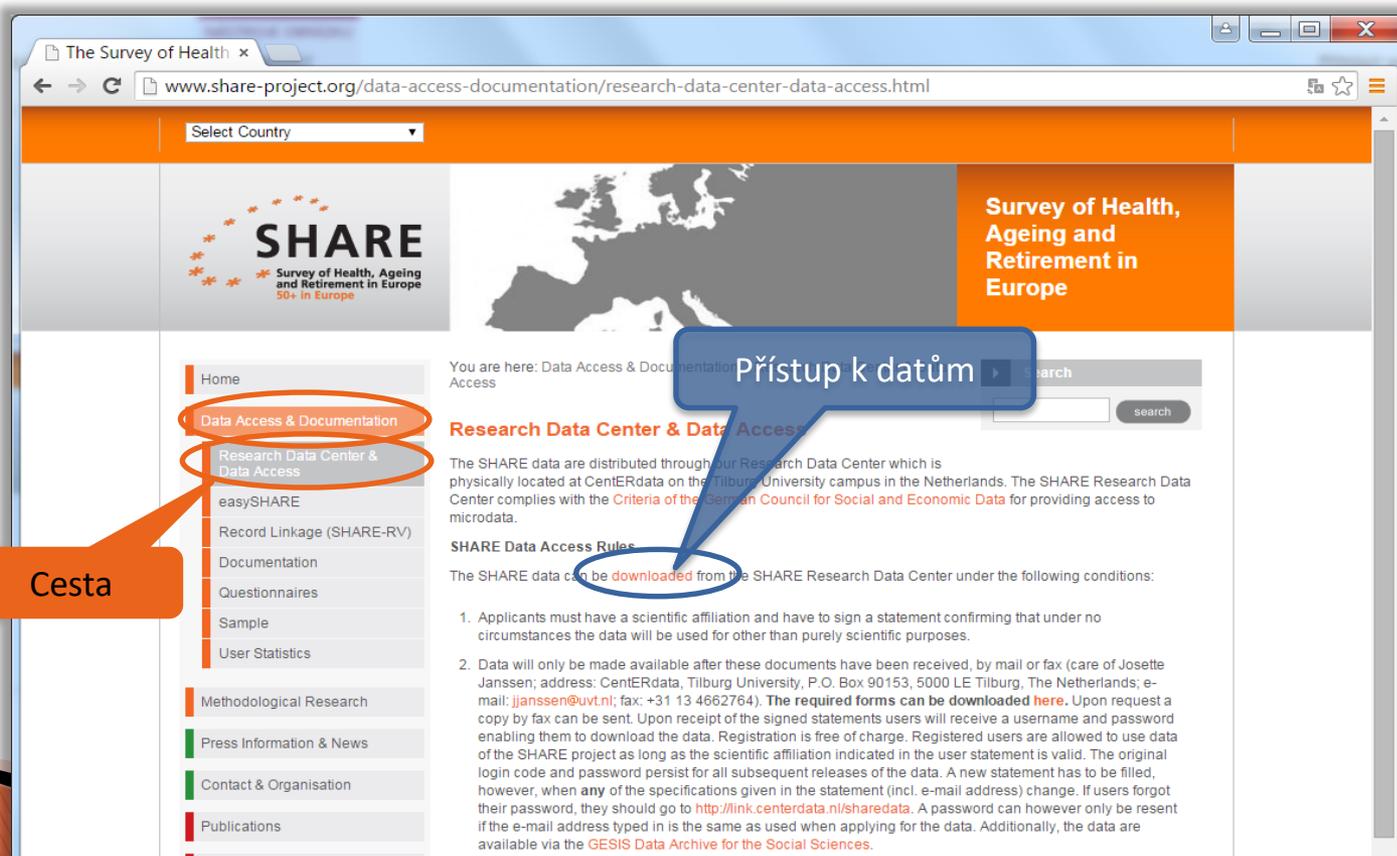
(e-mail address in capital letters)

(date) (place)

Please return this statement to: CentERdata, c/o Josette Janssen, P.O. Box 90153, 5000 LE Tilburg, NL; by mail to: jjanssen@uvt.nl or by fax to number: +31 13 466 2764.

Přístup k datům

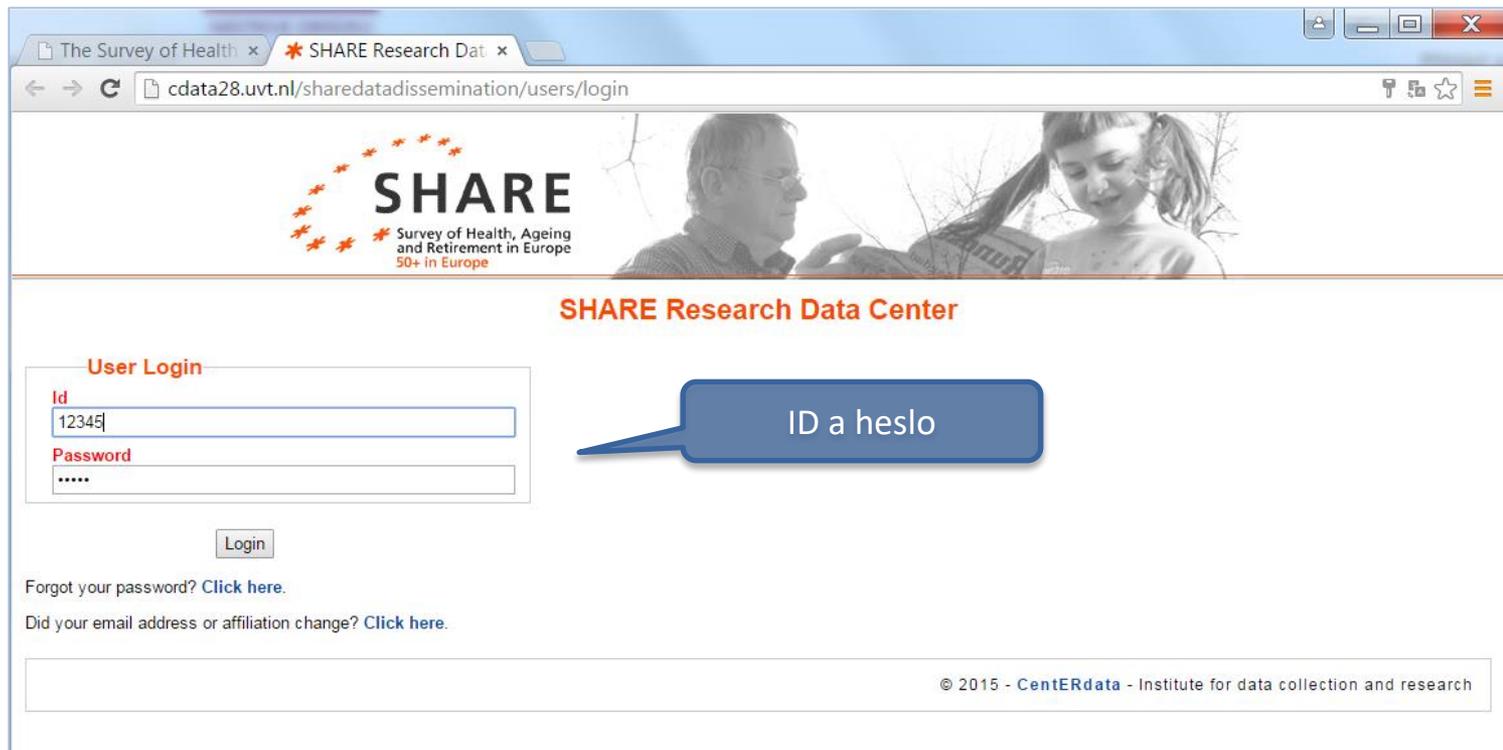
- ▶ Ze stránek <http://www.share-project.org> budete přeměrováni na stránky Data Center



The screenshot shows a web browser window displaying the SHARE website. The address bar shows the URL: www.share-project.org/data-access-documentation/research-data-center-data-access.html. The page features a navigation menu on the left with items: Home, Data Access & Documentation (circled in orange), Research Data Center & Data Access (circled in orange), easySHARE, Record Linkage (SHARE-RV), Documentation, Questionnaires, Sample, and User Statistics. A blue callout bubble labeled "Přístup k datům" points to the "Research Data Center & Data Access" link. An orange callout bubble labeled "Cesta" points to the "Data Access & Documentation" link. The main content area includes a "Select Country" dropdown, the SHARE logo, a map of Europe, and the text "Survey of Health, Ageing and Retirement in Europe". Below this, the page title is "Research Data Center & Data Access". The text states: "The SHARE data are distributed through our Research Data Center which is physically located at CentERdata on the Tilburg University campus in the Netherlands. The SHARE Research Data Center complies with the [Criteria of the European Council for Social and Economic Data](#) for providing access to microdata." Under the heading "SHARE Data Access Rules", it says: "The SHARE data can be downloaded from the SHARE Research Data Center under the following conditions:". A blue circle highlights the word "downloaded". The conditions listed are:

1. Applicants must have a scientific affiliation and have to sign a statement confirming that under no circumstances the data will be used for other than purely scientific purposes.
2. Data will only be made available after these documents have been received, by mail or fax (care of Josette Janssen; address: CentERdata, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands; e-mail: jjanssen@uvt.nl; fax: +31 13 4662764). **The required forms can be downloaded here.** Upon request a copy by fax can be sent. Upon receipt of the signed statements users will receive a username and password enabling them to download the data. Registration is free of charge. Registered users are allowed to use data of the SHARE project as long as the scientific affiliation indicated in the user statement is valid. The original login code and password persist for all subsequent releases of the data. A new statement has to be filled, however, when **any** of the specifications given in the statement (incl. e-mail address) change. If users forgot their password, they should go to <http://link.centerdata.nl/sharedata>. A password can however only be reset if the e-mail address typed in is the same as used when applying for the data. Additionally, the data are available via the [GESIS Data Archive for the Social Sciences](#).

Přístup k datům (login)



The screenshot shows a web browser window with the URL `cdata28.uvt.nl/sharedatadissemination/users/login`. The page features the SHARE logo and a banner image of an elderly man and a young girl. The main heading is "SHARE Research Data Center". Below it, there is a "User Login" section with two input fields: "Id" (containing "12345") and "Password" (containing "*****"). A "Login" button is positioned below these fields. A blue callout box with a pointer to the input fields contains the text "ID a heslo". At the bottom of the page, there are links for "Forgot your password? Click here." and "Did your email address or affiliation change? Click here." and a footer with the text "© 2015 - CentERdata - Institute for data collection and research".

SHARE
Survey of Health, Ageing
and Retirement in Europe
50+ in Europe

SHARE Research Data Center

User Login

Id
12345

Password

Login

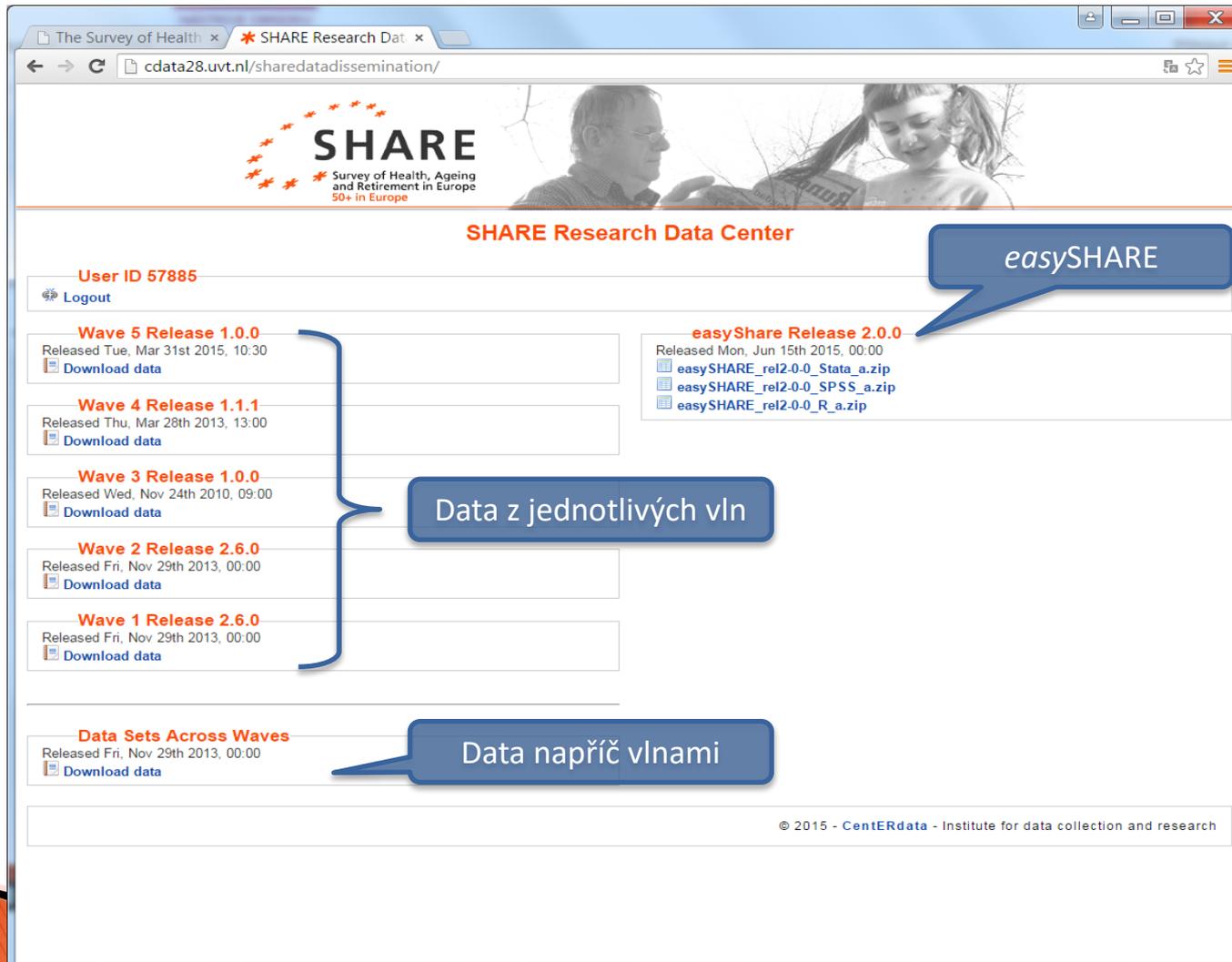
Forgot your password? [Click here.](#)

Did your email address or affiliation change? [Click here.](#)

© 2015 - CentERdata - Institute for data collection and research

ID a heslo

Úvodní obrazovka



The screenshot shows the SHARE Research Data Center website. At the top, there is a navigation bar with the SHARE logo and the text "Survey of Health, Ageing and Retirement in Europe 50+ in Europe". Below this, the page title "SHARE Research Data Center" is displayed. The main content area is divided into several sections:

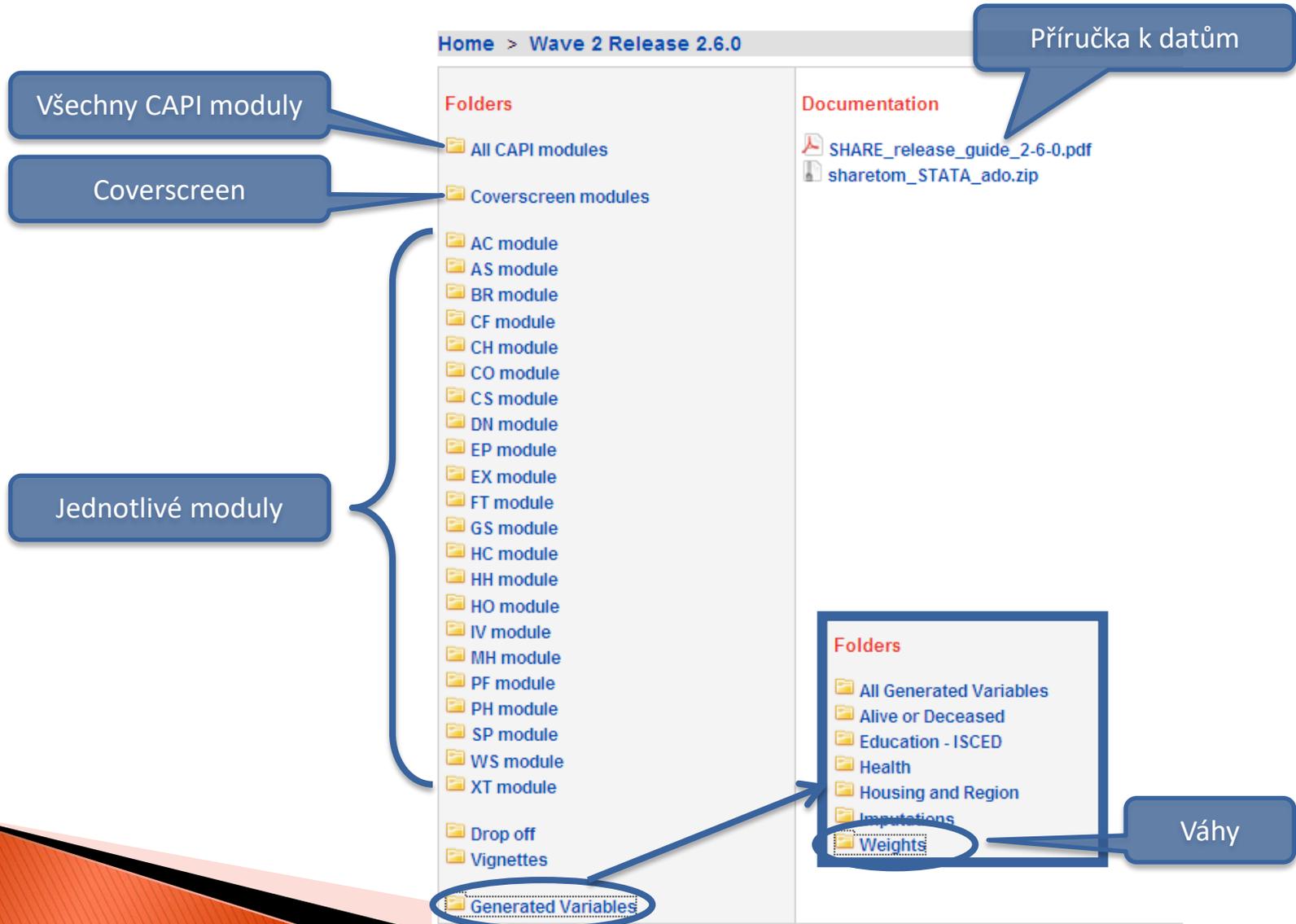
- User ID 57885**: A section with a "Logout" button.
- Wave 5 Release 1.0.0**: Released Tue, Mar 31st 2015, 10:30. Includes a "Download data" link.
- Wave 4 Release 1.1.1**: Released Thu, Mar 28th 2013, 13:00. Includes a "Download data" link.
- Wave 3 Release 1.0.0**: Released Wed, Nov 24th 2010, 09:00. Includes a "Download data" link.
- Wave 2 Release 2.6.0**: Released Fri, Nov 29th 2013, 00:00. Includes a "Download data" link.
- Wave 1 Release 2.6.0**: Released Fri, Nov 29th 2013, 00:00. Includes a "Download data" link.
- easyShare Release 2.0.0**: Released Mon, Jun 15th 2015, 00:00. Includes links for "easySHARE_rel2-0-0_Stata_a.zip", "easySHARE_rel2-0-0_SPSS_a.zip", and "easySHARE_rel2-0-0_R_a.zip".
- Data Sets Across Waves**: Released Fri, Nov 29th 2013, 00:00. Includes a "Download data" link.

Annotations on the page:

- A blue speech bubble labeled "easySHARE" points to the "easyShare Release 2.0.0" section.
- A blue bracket labeled "Data z jednotlivých vln" (Data from individual waves) encompasses the Wave 1 through Wave 5 release sections.
- A blue speech bubble labeled "Data napříč vlnami" (Data across waves) points to the "Data Sets Across Waves" section.

© 2015 - CentERdata - Institute for data collection and research

Struktura uložení dat

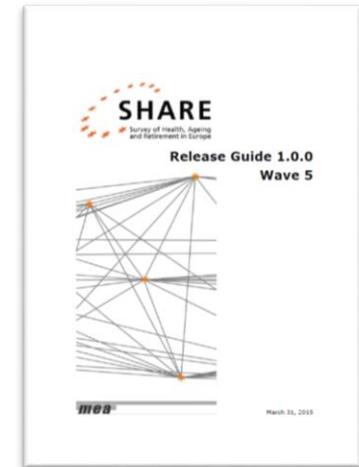


Každá publikace založená na datech SHARE **musí** obsahovat toto ujednání:

- ▶ *This paper uses data from SHARE Wave 5 release 1.0.0, as of March 31st 2015 (DOI: 10.6103/SHARE.w5.100) or SHARE Wave 4 release 1.1.1, as of March 28th 2013 (DOI: 10.6103/SHARE.w4.111) or SHARE Waves 1 and 2 release 2.6.0, as of November 29th 2013 (DOI: 10.6103/SHARE.w1.260 and 10.6103/SHARE.w2.260) or SHARELIFE release 1.0.0, as of November 24th 2010 (DOI: 10.6103/SHARE.w3.100). The SHARE data collection has been primarily funded by the European Commission through the 5th Framework Programme (project QLK6-CT-2001-00360 in the thematic programme Quality of Life), through the 6th Framework Programme (projects SHARE-I3, RII-CT-2006-062193, COMPARE, CIT5-CT-2005-028857, and SHARELIFE, CIT4-CT-2006-028812) and through the 7th Framework Programme (SHARE-PREP, N° 211909, SHARE-LEAP, N° 227822 and SHARE M4, N° 261982). Additional funding from the U.S. National Institute on Aging (U01 AG09740-13S2, P01 AG005842, P01 AG08291, P30 AG12815, R21 AG025169, Y1-AG-4553-01, IAG BSR06-11 and OGHA 04-064) and the German Ministry of Education and Research as well as from various national sources is gratefully acknowledged (see www.share-project.org for a full list of funding institutions)*

Co musíme mít před analýzou připraveno?

- ▶ Stažené datové soubory
 - ▶ Průvodce SHARE daty (zejm. vlna 1 & 2)
 - ▶ Dotazník
-
- ▶ Pokročilí uživatelé:
 - Innovations and Methodology (dané vlny)



Stažení dotazníku



Cesta

- Home
- Data Access & Documentation**
- Research Data Center & Data Access
- easySHARE
- Record Linkage (SHARE-RV)
- Documentation
- Questionnaires**
- Questionnaire Wave 1
- Questionnaire Wave 2
- Questionnaire Wave 3 (SHARELIFE)
- Questionnaire Wave 4
- Questionnaire Wave 5**
- Sample
- User Statistics
- Methodological Research
- Press Information & News
- Contact & Organisation
- Publications
- FAQs

You are here: Data Access & Documentation · Questionnaires · Questionnaire Wave 5

Search search

Questionnaire Wave 5

	CAPI Coverscreen	CAPI main questionnaire	Showcards	End of life questionnaire*	Drop-off questionnaire
English generic questionnaire Wave 5	X	X	X		
Austria	X	X	X		X
Belgium (Dutch)	X	X	X		
Belgium (French)	X	X	X		
Czech Republic	X	X	X		X (A) - X (B)
Denmark	X	X	X		
Estonia (Estonian)	X		X		
Estonia (Russian)	X		X		
France	X	X	X		
Germany	X	X	X		
Israel					X
Israel (Arabic)		X	X		
Israel (Hebrew)	X	X	X		
Israel (Russian)	X	X	X		
Italy	X	X	X		
Luxemburg (French)	X	X	X		
Luxemburg (German)	X	X	X		
Netherlands	X	X			
Slovenia	X	X	X		

CZ hlavní dotazník w5

CZ DROP-OFF dotazník w5

Důležité - než začneme se SHARE

- 1. Jakou informaci chceme z dat získat?**
 - Téma našeho analytického problému (výběr modulů)
- 2. Co bude naše základní jednotka (jeden případ), příp. jaké bude mít specifické vlastnosti (věk, pohlaví aj.)?**
 - Respondent, respondent + partner, či domácnost, příp. ženy mezi 50 a 65 lety, důchodci starší 70 let
- 3. Jaká bude naše základní územní jednotka?**
 - Jeden stát, celá oblast (např. Skandinávie), nebo porovnání několika (všech) států (či oblastí) v rámci Evropy; lze i nižší územní celky (NUTS 2, kraje) či dle velikosti obce
- 4. Chceme analyzovat určitý stav, porovnat ho se stavem před 5 lety (tzv. *cross-sectional*), nebo se podívat se na jeho vývoj v čase v rámci panelu (tzv. *longitudinal*)?**
 - Jedna vlna, několik spojených vln kohortně nebo longitudinálně svázaných (panel)

⇒ Podle toho data zvážíme



Zjednodušení

1. tematická oblast, resp. modul
 2. základní jednotka (jeden případ)
 3. územní jednotka
 4. čas, resp. vlna
-
- ▶ díky obsáhlému datovému souboru lze kombinovat a filtrovat jednoduše i složitěji, dat bude dostatek

Proč je nutné znát předem směr a způsob zkoumání?

- ▶ data z výzkumu jsou **velmi obsáhlá a komplexní**, při spojení všech modulů dané vlny (případně napojení další vln) vznikne obrovská datamatice
 - např. při spojení vln 2, 3 a 4 se jedná o 7500 proměnných a 86 300 řádků/dotazníků, nehledě na samotnou proceduru napojování (nutno spojit cca 100 souborů)
- ▶ doporučená výzkumná strategie:
od jedné tematické části k většímu celku



Data *easy*SHARE

»» struktura dat, proměnné

O datech *easy*SHARE (1)

- ▶ zjednodušená verze dat výzkumu SHARE vytvořená především k studijním účelům a k lepšímu seznámení s komplexním datovým souborem
- ▶ zahrnuje 4 proběhnuté vlny výzkumu
- ▶ přístup k datům má každý vědec, který data použije čistě k vědeckým účelům
- ▶ **základní jednotka:** jeden respondent v jedné vlně

O datech *easy*SHARE (2)

- ▶ Data vybrána primárně pro možnost srovnání s HRS
- ▶ Pokrývají všechny moduly SHARE
- ▶ Výběr dat podle pravidel:
 - Snaha o co nejnižší úroveň chybějících informací
 - Možnost mezinárodního srovnání
 - Maximální možnost srovnání mezi jednotlivými vlnami
 - (tj. longitudinální ukazatele)
 - Indexy místo proměnných
 - Žádné složité filtry

Struktura easySHARE dat (1)

1. SOCIO-DEMOGRAFICKÉ PROMĚNNÉ

- age, gender, country of birth, citizenship, education, religion, marital status, age and gender of partner

2. DOMÁCNOST A JEJÍ SLOŽENÍ

- living with partner in the same household, household size, children living in the household

3. SOCIÁLNÍ PODPORA A SOCIÁLNÍ SÍŤ

- mother/father alive, number of children, residential proximity of children, number of grandchildren, number of living siblings, social activities, received and given social support

4. DĚTSTVÍ

- number of books at age ten, occupation of main breadwinner at age ten, relative mathematical skills at age ten, relative language skills at age ten

Struktura easySHARE dat (2)

5. ZDRAVÍ

- self-perceived health, number of chronic diseases, mental health variables, depression scale EURO-D, CASP-12 index for quality of life and well-being, health care utilization, grip strength, body mass index, smoking and drinking behavior, vigorous activities/sports

6. OMEZENÍ V ŽIVOTĚ – INDEXY

- mobility index, large muscle index, activities of daily living index, gross motor skills index, fine motor skills index, instrumental activities of daily living index, cognitive functions

7. PRÁCE A ČAS

- current job situation, term of main job, working hours per week, satisfaction with main job, early retirement plans, able to make ends meet, monthly expenditure on food, wave-specific household income percentiles

POZN: více k proměnným
v brožuře *Guide to easySHARE*

Základní jednotka

- ▶ každý řádek představuje 1 respondenta v jedné vlně

Table 1: Data structure

Mergeid	wave	country	wavepart
AT-986403-01	1	Austria	1234
AT-986403-01	2	Austria	1234
AT-986403-01	3	Austria	1234
AT-986403-01	4	Austria	1234
SE-209636-01	1	Sweden	14
SE-209636-01	4	Sweden	14

identifikátor
respondenta

vlna

účast ve
vlnách

Missing codes

- ▶ chybějící hodnoty v datech *easySHARE* jsou různého typu

Hodnota	Popis hodnoty
-1	don't know
-2	refusal
-3	implausible value/suspected wrong
-7	not yet coded
-9	filtered: did not go to school
-12	don't know / refusal
-13	not asked in this wave
-14	not asked in this country
-15	no information
-16	no drop-off (information in drop-off in this wave)

Příprava dat na analýzu *easySHARE*



Doporučená literatura

- ▶ **Mareš, Rabušic, Soukup.** 2015. *Analýza sociálněvědních dat (nejen) v SPSS.* Brno: Masarykova univerzita.
- ▶ **Pakosta, Fučík.** 2009. Vybrané metody analýzy panelových dat. *Data a výzkum – SDA Info*, Vol. 3, No. 1: 77-96.
- ▶ **Kalvas, Kreidl, Váně, Štípková.** 2009. Modelování panelových dat s dichotomickou závisle proměnnou: obecné principy a ilustrace v programu STATA. *Data a výzkum – SDA Info*, Vol.3, No.1: 96-123.
- ▶ Celé monočíslo o práci s panelovými daty

Datový soubor easySHARE

*easySHARE_rel5-0-0.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Extensions Window Help

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	R
1	mergeid	String	12	0	Person identifie...	None	None	12	Left	Nominal	Input
2	hhid	String	11	0	Household iden...	None	None	11	Left	Nominal	Input
3	coupleid	String	15	0	Couple identifie...	None	None	15	Left	Nominal	Input
4	wave	Numeric	8	0	Wave	None	None	8	Right	Nominal	Input
5	wavepart	Numeric	8	0	Wave participat...	None	None	8	Right	Scale	Input
6	int_year	Numeric	8	0	Interview year	{-15, -15. no...	None	8	Right	Scale	Input
7	int_month	Numeric	8	0	Interview month	{-15, -15. no...	None	8	Right	Nominal	Input
8	country	Numeric	8	0	Country identifier	{-15, -15. no...	None	8	Right	Scale	Input
9	country_mod	Numeric	8	0	Country identifi...	{-15, -15. no...	None	8	Right	Scale	Input
10	language	Numeric	8	0	Language of qu...	{-15, -15. no...	None	8	Right	Scale	Input
11	female	Numeric	8	0	Gender: female...	{-15, -15. no...	None	8	Right	Nominal	Input
12	dn002_mod	Numeric	8	0	Month of birth	{-15, -15. no...	None	8	Right	Scale	Input
13	dn003_mod	Numeric	8	0	Year of birth	{-15, -15. no...	None	8	Right	Scale	Input
14	dn004_mod	Numeric	37	0	... in the cou...	{-15, -15. no...	None	8	Right	Scale	Input
15	age	Numeric	40	0		{-15, -15. no...	None	8	Right	Scale	Input
16	birth_country	Numeric	8	0		{-15, -15. no...	None	8	Right	Scale	Input
17	citizenship	Numeric	8	0		{-15, -15. no...	None	8	Right	Scale	Input

Data View Variable View

IBM SPSS Statistics Processor is ready Unicode:ON

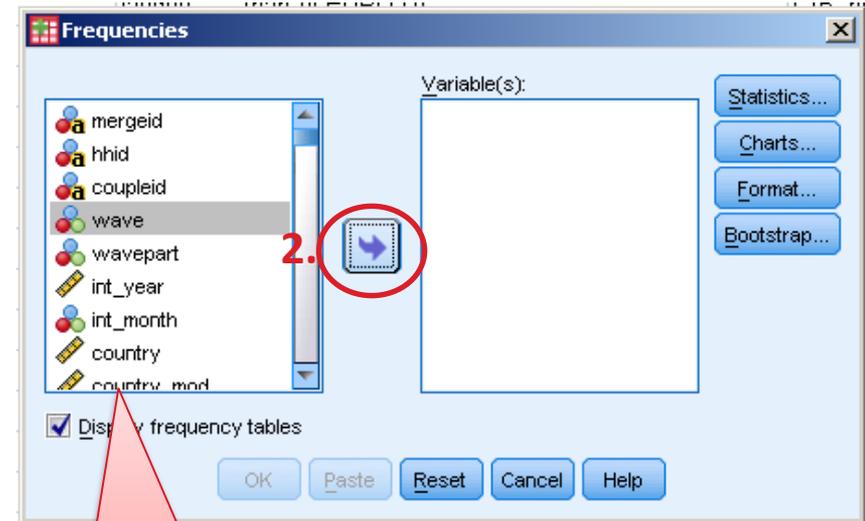
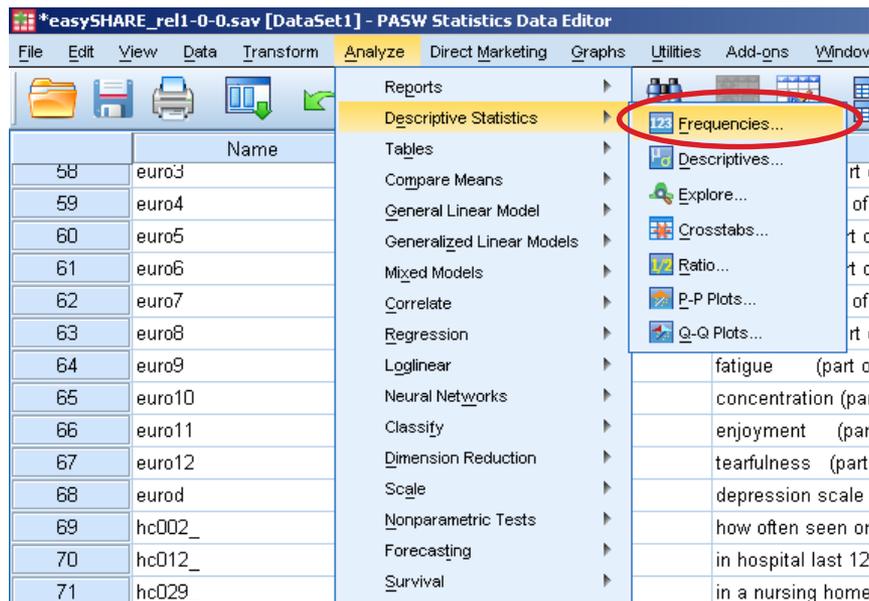
přehled proměnných

datová matice

FREKVENCE

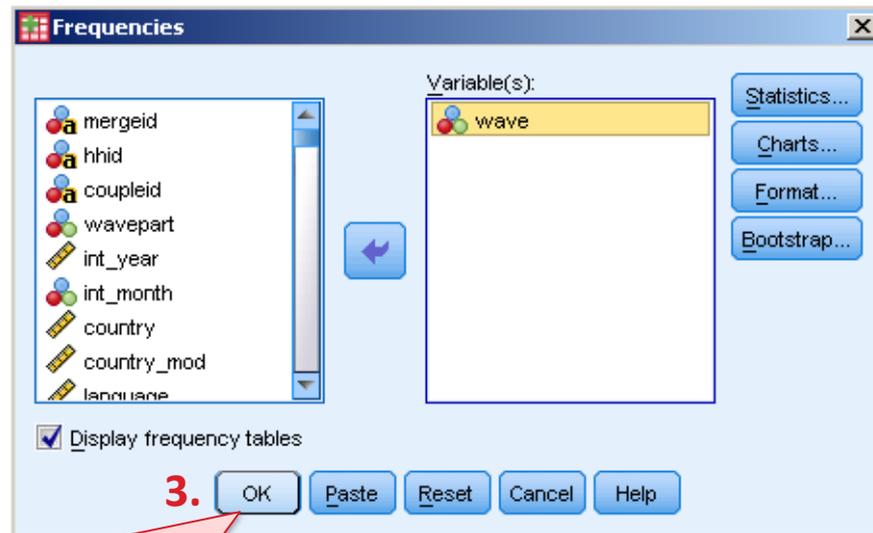
ÚKOL – počet respondentů
v jednotlivých vlnách (*wave*)

- ▶ neboli třídění prvního stupně – zobrazí četnosti hodnot v rámci jedné proměnné



1. Vybrat
proměnnou/proměnné

FREKVENCE (2)



3. Buď **OK** (provede se příkaz), nebo **Paste** (příkaz se zobrazí v novém souboru - syntaxi – pro dokončení nutné spustit syntax)

- ▶ `stx: FREQUENCIES wave.` (nebo stačí jen `FRE wave.`)

FREKVENCE (3) - výstup

The screenshot shows the PASW Statistics Viewer interface. The main window displays the output of a 'Frequencies' procedure. The data source is '[DataSet1] E:\CERGEI_práce\easySHARE\easySHARE_re11-0-0.sav'. The output is divided into two sections: 'Statistics' and a table of frequencies.

Statistics

Statistic	Valid	Missing
N	152871	0

Table of Frequencies

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	30816	20,2	20,2	20,2
Valid 2	36730	24,0	24,0	44,2
Valid 3	26836	17,6	17,6	61,7
Valid 4	58489	38,3	38,3	100,0
Total	152871	100,0	100,0	

Two callout boxes provide additional context:

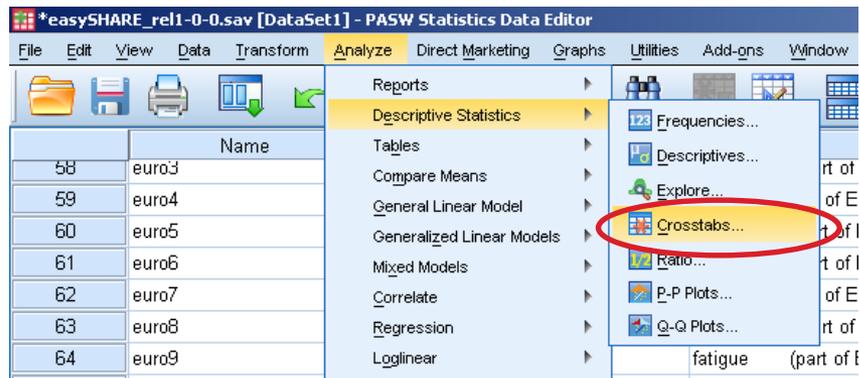
- An orange callout box points to the 'Statistics' table with the text: "Přehled platných a chybějících hodnot" (Overview of valid and missing values).
- A larger orange callout box points to the frequency table with the text: "Tabulka absolutních, relativních a kumulativních četností" (Table of absolute, relative, and cumulative frequencies).

The status bar at the bottom indicates "PASW Statistics Processor is ready" and "H: 4.19, W: 5.83 in".

CROSSTABS (1)

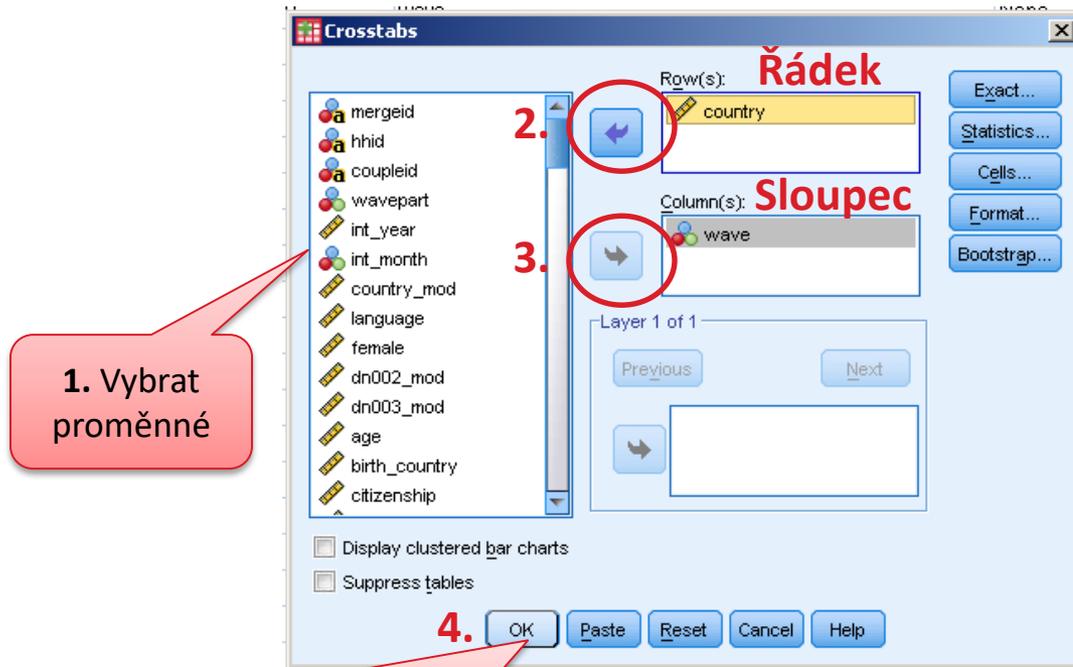
ÚKOL – počet respondentů
v jednotlivých vlnách (*wave*)
v jednotlivých zemích (*country*)

- ▶ neboli třídění druhého stupně – „křížení“ dvou (příp. více) proměnných



- ▶ **DŮLEŽITÉ** – nutné mít správně nadefinované missing values – do výsledné tabulky vstoupí jen hodnoty od respondentů, kteří nemají ani v jedné ze zkoumaných hodnot missing value

CROSSTABS (2)



1. Vybrat proměnné

4. Buď **OK** (provede se příkaz), nebo **Paste** (příkaz se zobrazí v syntaxi – pro dokončení nutné spustit syntax)

► stx: CROSSTABS country BY wave. (také jen CRO...)

CROSSTABS (3) - výstup

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
country country identifier *	152871	100,0%	0	,0%	152871	100,0%
wave wave						

country country identifier * wave wave Crosstabulation

Count

		wave wave				Total
		1	2	3	4	
country country identifier	11 Austria	1594	1192	847	5286	8919
	12 Germany	3008	2568	1852	1572	9000
	13 Sweden	3053	2745	1893	1951	9642
	14 Netherlands	2979	2661	2210	2762	10612
	15 Spain	2396	2228	2048	3570	10242
	16 Italy	2559	2983	2492	3583	11617
	17 France	3193	2968	2483	5857	14501
	18 Denmark	1707	2616	2141	2276	8740
	19 Greece	2898	3243	2951	0	9092
	20 Switzerland	1004	1462	1296	3750	7512
	23 Belgium	3827	3169	2832	5300	15128
	25 Israel	2598	2464	0	0	5062
	28 Czechia	0	2830	1873	6118	10821

Přehledová
tabulka počtu
případů
vstupujících do
crosstabs

Výstupní tabulka

RECODE INTO - zadání

ÚKOL – z proměnné *recall_2* vytvořit novou proměnnou *pamet* se třemi kategoriemi

recall_2 recall of words, delayed (based on cf016tot)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	12761	8,3	10,4	10,4
	1	8940	5,8	7,3	17,6
	2	14498	9,5	11,8	29,4
	3	21913	14,3	17,8	47,3
	4	23528	15,4	19,1	66,4
	5	19223	12,6	15,6	82,0
	6	12053	7,9	9,8	91,8
	7	6031	3,9	4,9	96,7
	8	2555	1,7	2,1	98,8
	9	1042	,7	,8	99,6
	10	436	,3	,4	100,0
	Total	122980	80,4	100,0	
Missing	-15 no information	3055	2,0		
	-13 not asked in this wave	26836	17,6		
	Total	29891	19,6		
	Total	152871	100,0		

1 špatná

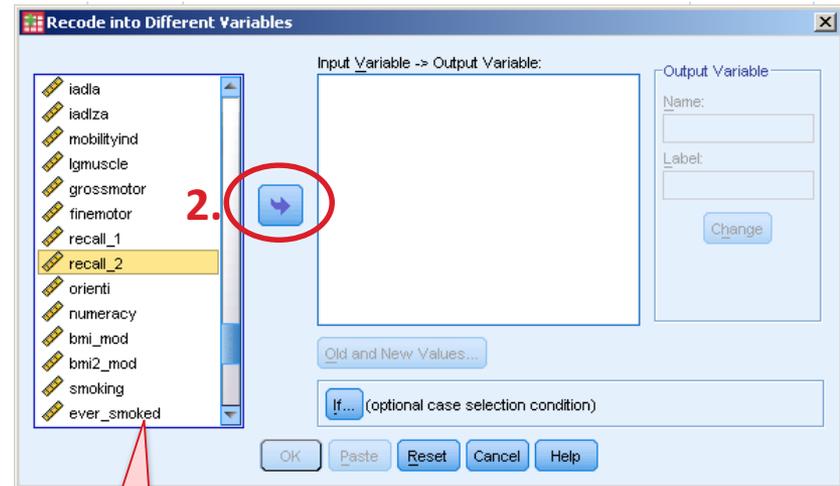
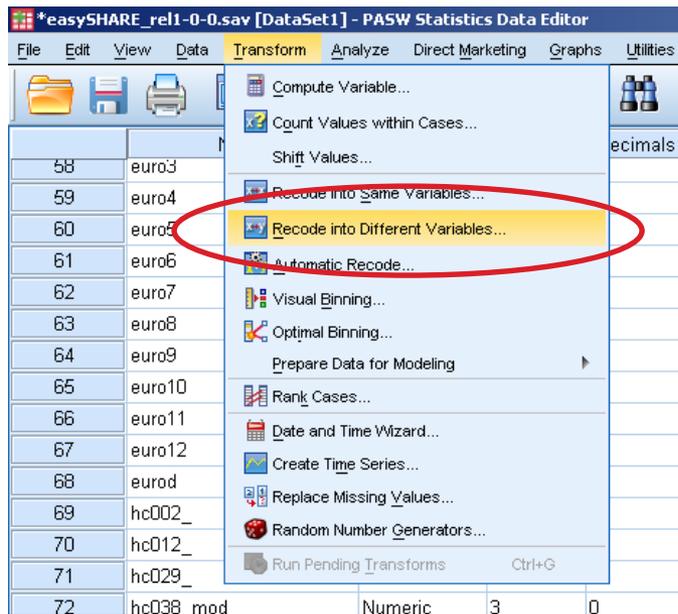
2 průměrná

3 výborná

Missing

RECODE INTO (1)

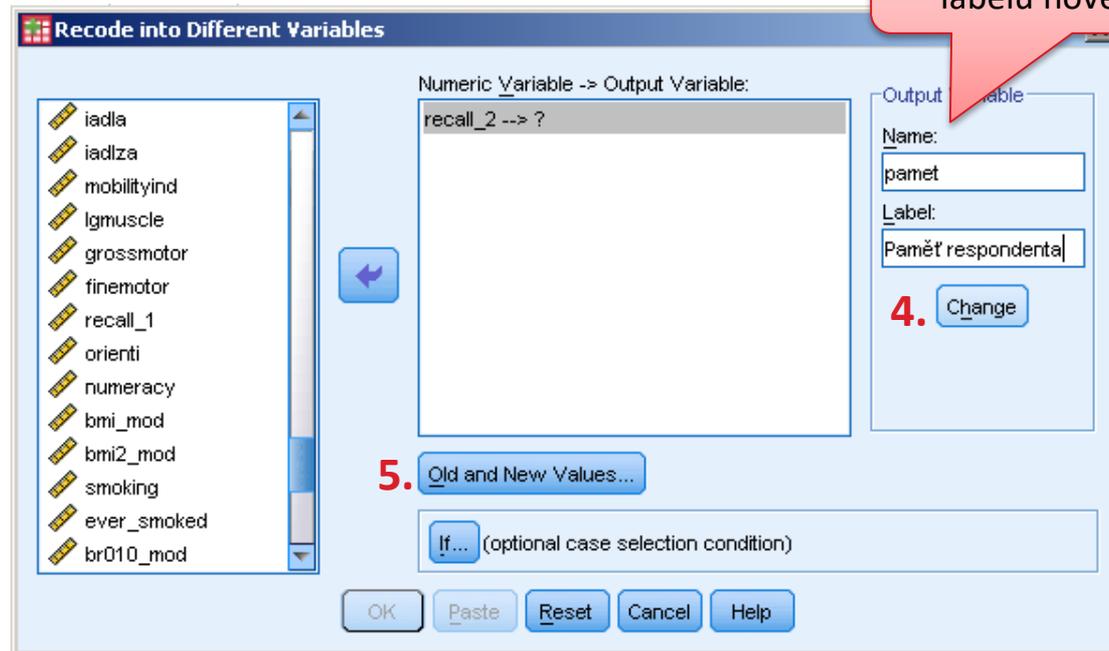
- ▶ mění hodnoty proměnné na jiné a vepíše je do nové proměnné



1. vybrat
proměnnou

RECODE INTO (2)

3. zadání jména, případně
labelu nové proměnné



RECODE INTO (3)

Navolím změny hodnot

zadávání jednotlivě

missing

rozsah

menší než X

od X výš

všechny ostatní

Old Value **6. Old Value**

Value:

System-missing

System- or user-missing

Range:

through

Range, LOWEST through value:

Range, value through HIGHEST:

7

All other values

New Value **7. New value**

Value: 3

System-missing

Copy old value(s)

8. Old --> New:

0 --> 1

SYSMIS --> SYSMIS

1 thru 2 --> 1

3 thru 6 --> 2

7 thru Highest --> 3

Add

Change

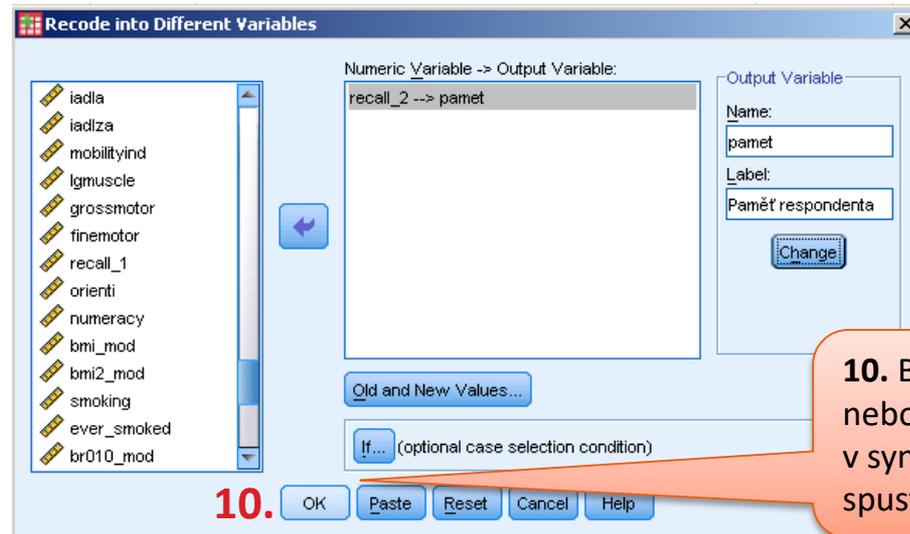
Remove

Output variables are strings Width: 8

Convert numeric strings to numbers ('5' -> 5)

9. Continue Cancel Help

RECODE INTO (4)



- ▶ stx: RECODE recall_2 (0=1) (1 thru 2=1) (3 thru 6=2) (7 thru Highest=3) (SYSMIS=SYSMIS) INTO pamet.
VARIABLE LABELS pamet 'Paměť respondenta'.
EXECUTE.

RECODE INTO (5) - kontrola

- ▶ pro kontrolu správného překódování porovnáme v kontingenční tabulce starou a novou proměnnou

recall_2 recall of words, delayed (based on cf016tot) * pamet Paměť respondentů
Crosstabulation

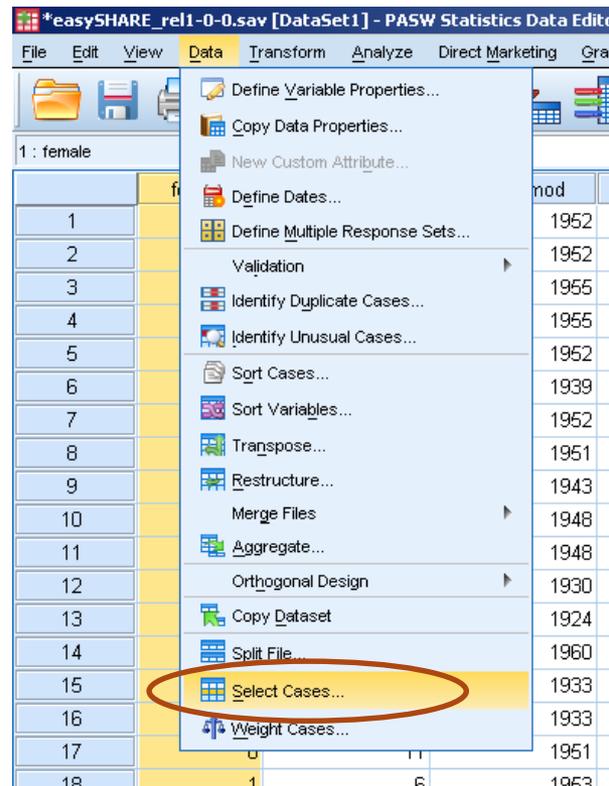
Count

		pamet Paměť respondentů			Total
		1,00 špatná paměť	2,00 průměrná paměť	3,00 výborná paměť	
recall_2 recall of words, delayed (based on cf016tot)	0	12761	0	0	12761
	1	8940	0	0	8940
	2	14498	0	0	14498
	3	0	21913	0	21913
	4	0	23528	0	23528
	5	0	19223	0	19223
	6	0	12053	0	12053
	7	0	0	6031	6031
	8	0	0	2555	2555
	9	0	0	1042	1042
	10	0	0	436	436
Total		36199	76717	10064	122980

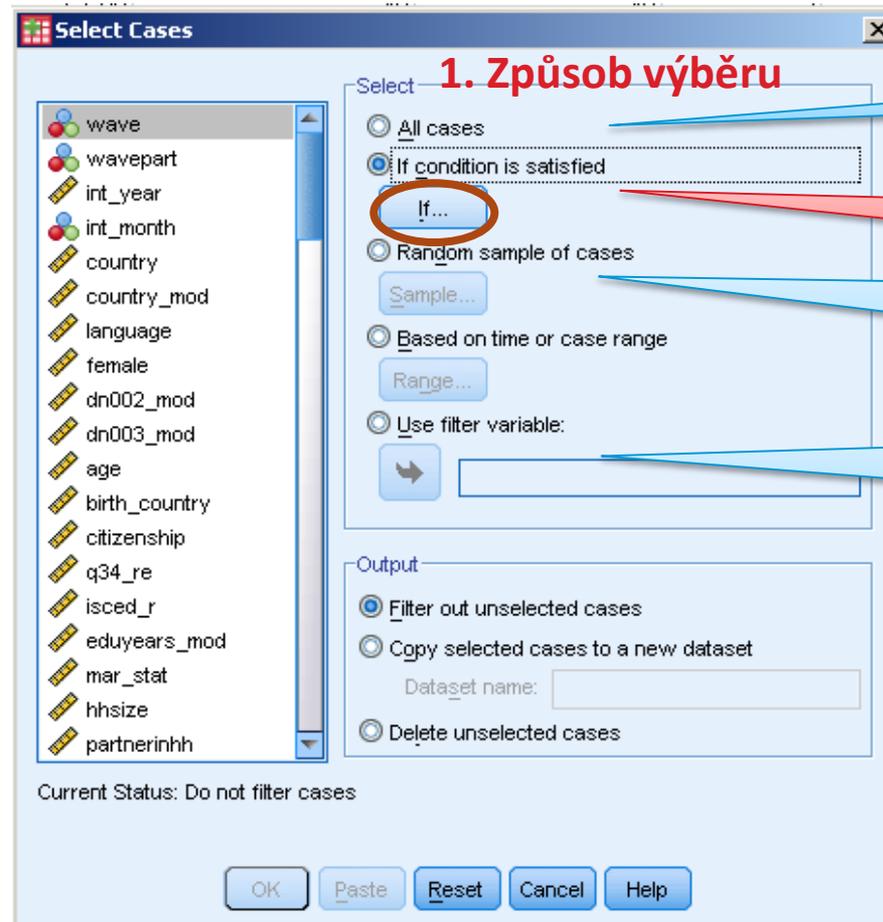
SELECT CASES (1)

ÚKOL – vyfiltrovat vlnu 4 (*wave*)

- ▶ vytvoření datového filtru (pracujeme s případy vybranými podle kritéria, ostatní si „skryjeme“ nebo odstraníme)



SELECT CASES (2)



1. Způsob výběru

Všechny případy

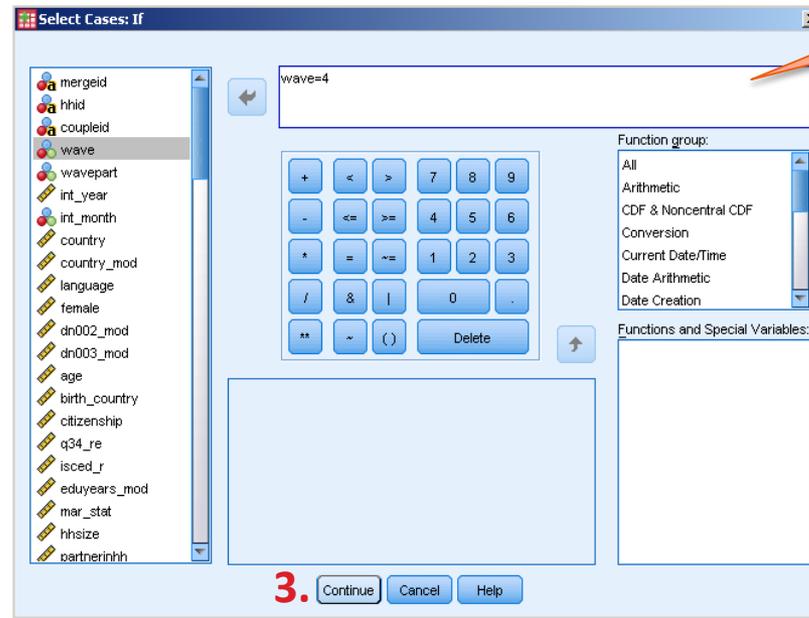
Vybere na základě podmínky

Náhodně vybere určitý počet či poměr případů

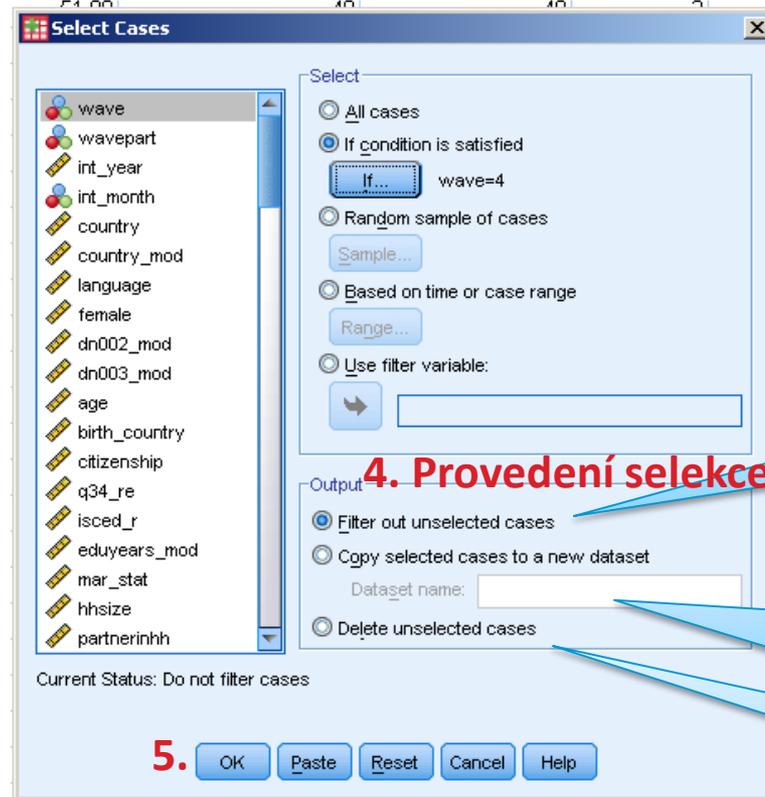
Použití filtrovací proměnné [0,1]

SELECT CASES (3) - podmínka

2. Stanovení podmínky



SELECT CASES (4) - selekce



4. Provedení selekce

A) Filtr na vybrané (nevybrané skryty)

B) Vytvoření nového datového souboru s vybranými daty – nutno zadat název datasetu

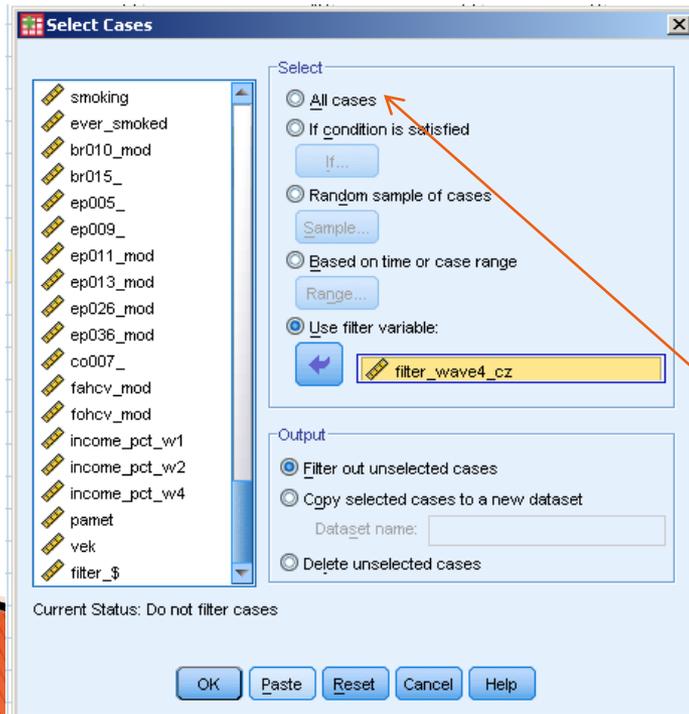
C) Vymazání nevybraných

SELECT CASES (5) - syntax

- ▶ stx A): USE ALL.
COMPUTE filter_\$(wave >= 2).
VARIABLE LABEL filter_\$ 'wave >= 2 (FILTER)'.
VALUE LABELS filter_\$ 0 'Not Selected' 1 'Selected'.
FORMAT filter_\$ (f1.0).
FILTER BY filter_\$.
EXECUTE.
- ▶ stx B): DATASET COPY vlna4.
DATASET ACTIVATE vlna4.
FILTER OFF.
USE ALL.
SELECT IF (wave=4).
EXECUTE.
- ▶ stx C): FILTER OFF.
USE ALL.
SELECT IF (wave=4).
EXECUTE.

SELECT CASES (6) - tip

- ▶ u filtrů založených na složitějších podmínkách či filtrů, které se budou používat v analýze opakovaně, je dobré vytvořit si proměnnou [0,1], kterou v případě potřeby spustíme jako filtr (způsob A)



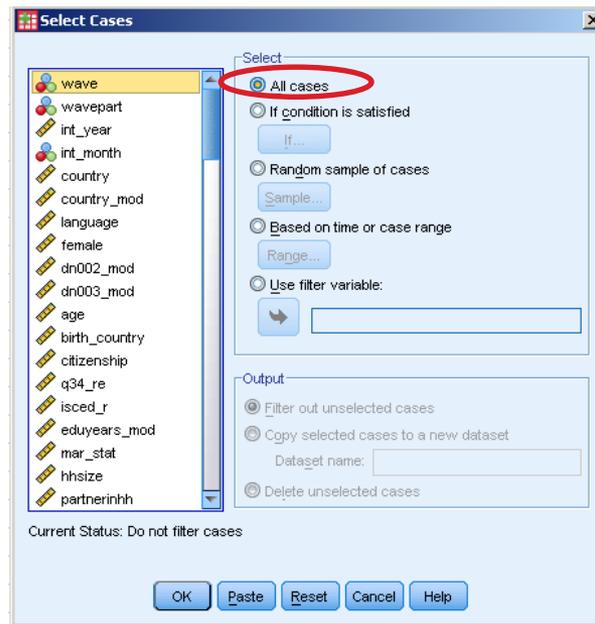
- ▶ stx: FILTER BY filter_wave4_cz.
EXE.

- ▶ Pozn. pro zrušení filtru All cases.

- ▶ stx: FILTER OFF.
EXE.

SELECT CASES (7) – zrušení filtru

- ▶ Pro zrušení filtru, tzn. zahrnutí všech případů do analýzy

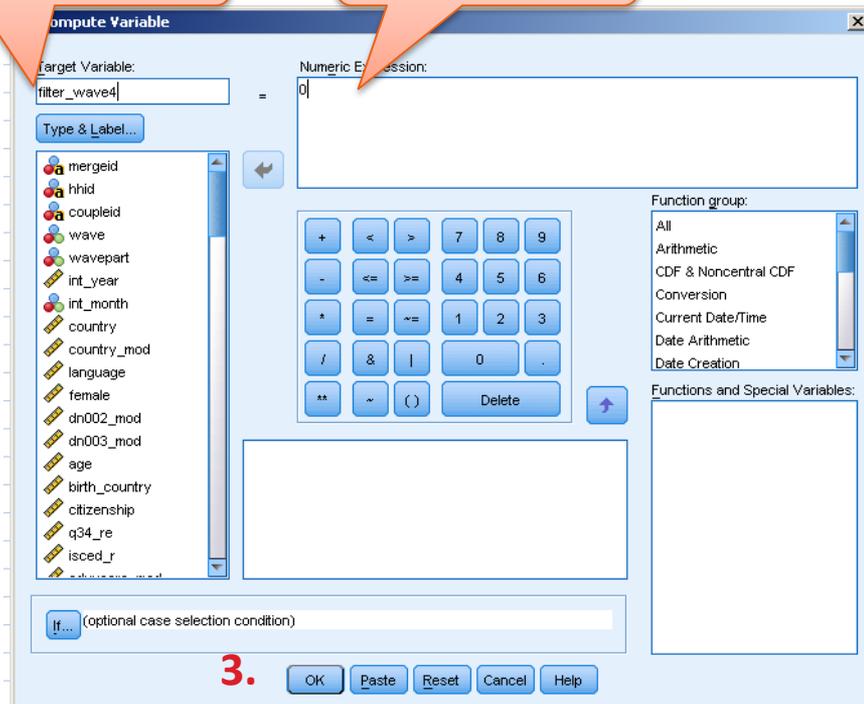


- ▶ stx: FILTER OFF.
EXE.

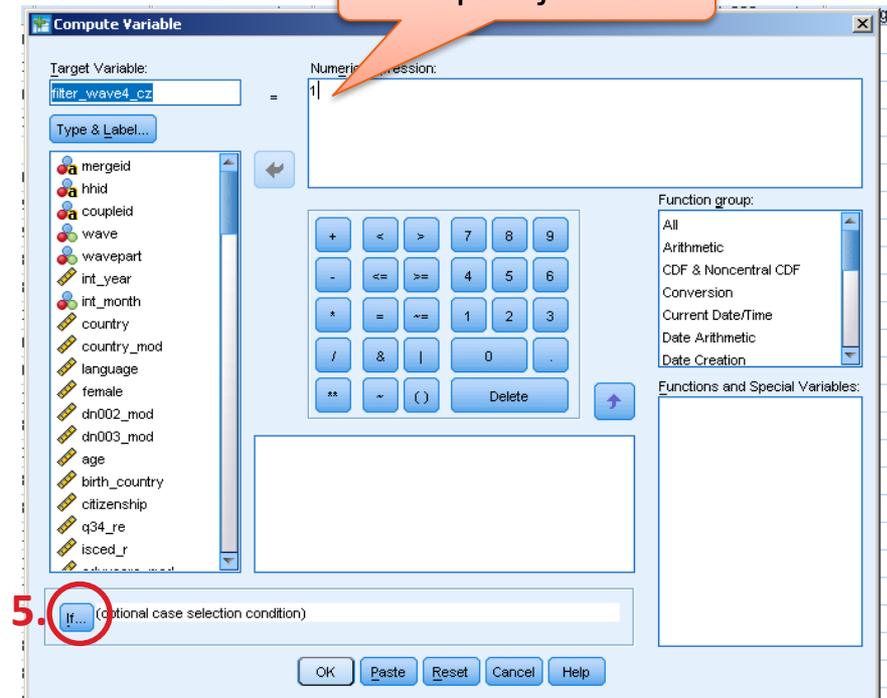
VÝPOČET FILTRU (1)

1. název filtru

2. napsat nulu

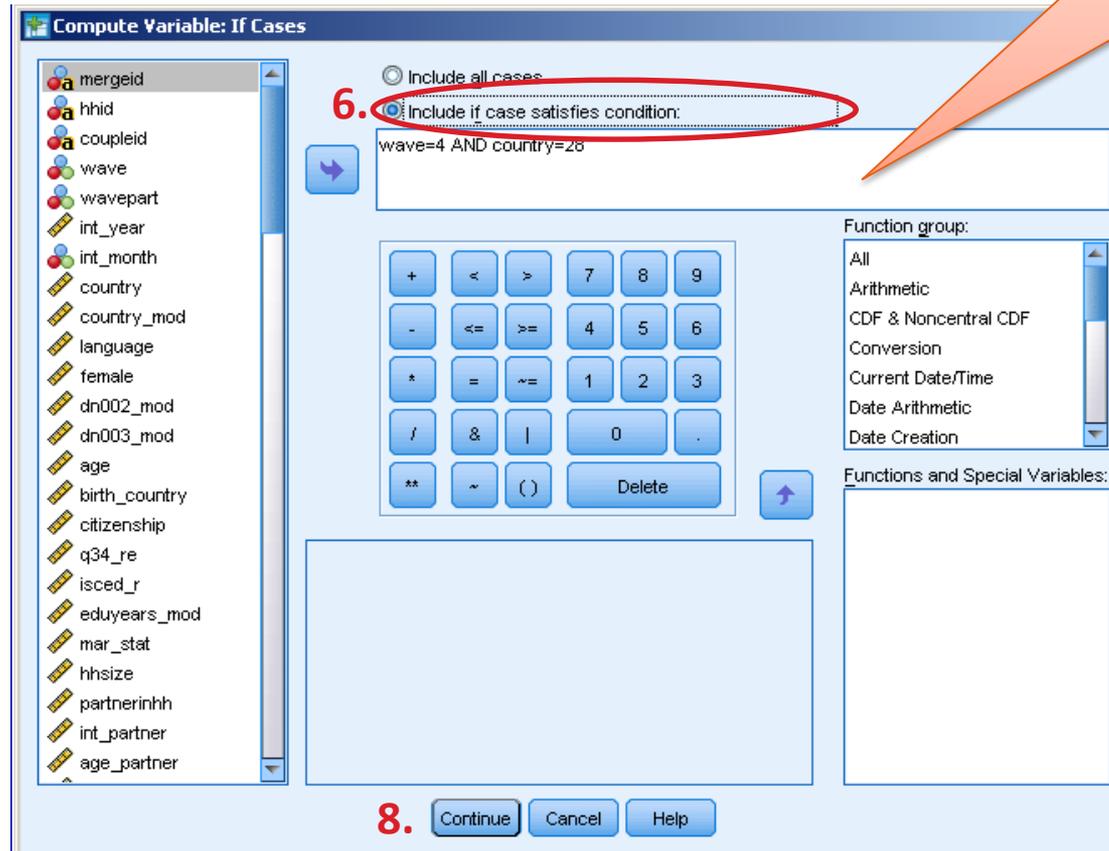


4. napsat jedničku

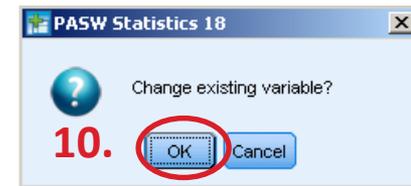
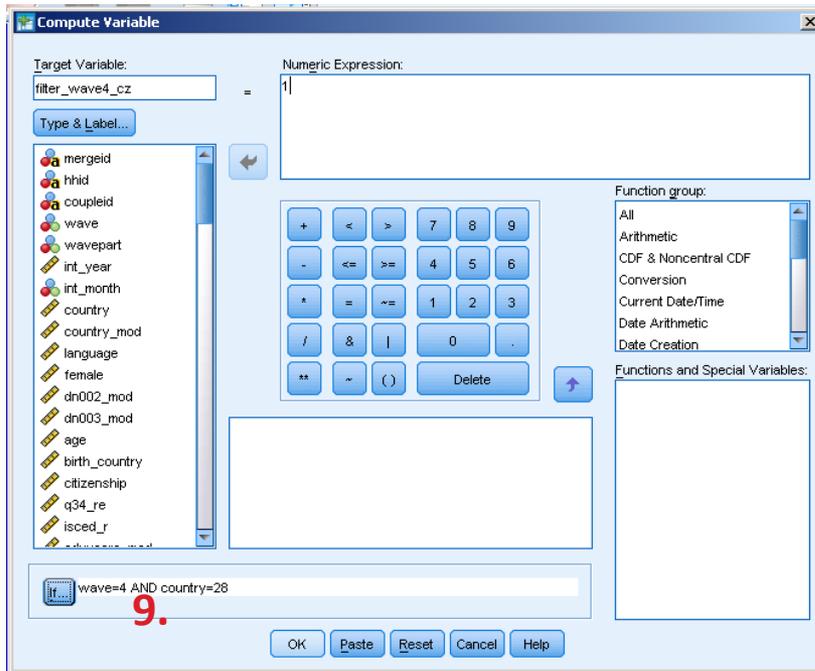


VÝPOČET FILTRU (2)

7. Logická podmínka
wave=4 AND country=28



VÝPOČET FILTRU (3)



- ▶ stx: COMPUTE filter_wave4_cz=0.
IF wave=4 AND country=28 filter_wave4_cz=1.
EXE.

SYNTAX TVORBY FILTRŮ

► COMPUTE filter_wave4_cz=0.
IF wave=4 AND country=28 filter_wave4_cz=1.
EXE.

Vypočítej proměnnou filter a přiřaď jí hodnotu 0.

Proved'.

Pokud případ splňuje **podmínku**, filtru přiřaď hodnotu 1.

LOGICKÉ OPERÁTORY V SLOŽENÝCH PODMÍNKÁCH (2 logické argumenty)

- ▶ AND – vyhodnotí jako 1, pokud jsou obě (všechny) podmínky jsou splněny
 - *IF wave=4 AND female=1* vybere všechny případy žen z vlny 4
- ▶ OR – vyhodnotí jako 1, pokud je splněna alespoň jedna podmínka
 - *IF wave=4 AND female=1* vybere všechny případy žen ze všech vln + všechny případy vlny 4
- ▶ NOT – negace, také v kombinaci s AND nebo OR
 - *IF not(wave=4)* vybere všechny případy kromě vlny 4

Jednoduché analýzy v datech *easy*SHARE



Výzkumná otázka 1

- ▶ Ekonomické důsledky mateřství na výši důchodu?
- ▶ POSTUP ?
- ▶ Výběr úrovně – které země, která vlna?
- ▶ Výběr proměnných (filtr pohlaví, mateřství; výše příjmu)
- ▶ Na datech easySHARE není tato otázka řešitelná, ukážeme si pouze test míry asociace statutu důchodkyně s dětmi a výší příjmu domácnosti

MISSING VALUES

- ▶ tzv. chybějící hodnoty
- ▶ obvykle hodnoty 99, někdy 0, - 1 apod.
- ▶ **vznik:** otázka respondentovi nepoložena (filtr), respondent odmítl odpovědět/nevěděl, odpověď nebyla uznána jako platná („nesmyslná“ hodnota)
- ▶ **Nutné nadefinovat před analýzou!**

Nadefinování missing values

The screenshot shows the PASW Statistics Data Editor interface. The main window displays a data table with columns: Name, Type, Width, Decimals, Label, Values, Missing, and Columns. The 'Missing' column for the variable 'gender_part...' is highlighted in yellow and contains the text '-15 -1, 5'. A red circle is drawn around this text. The 'Missing Values' dialog box is open, showing three radio button options: 'No missing values', 'Discrete missing values', and 'Range plus one optional discrete missing value'. The 'Range plus one optional discrete missing value' option is selected. The 'Low' field is set to '-15' and the 'High' field is set to '-1'. The 'Discrete value' field is empty. The 'OK' button is highlighted with a red box and a red arrow pointing to it.

bez chybějících hodnot

jednotlivé chybějící hodnoty

škála chybějících hodnot (od do) + možnost jednotlivé chybějící hodnoty

- ▶ stx: MISSING VALUES female ().
MISSING VALUES female (-15, -7, -3).
MISSING VALUES female (-15 thru -1).
- ▶ **TIP** - nadefinování stejných missing values pro všechny numerické proměnné
MISSING VALUES all (-15 thru -1).

Frekvence

- ▶ `stx:`
`FREQUENCIES`
`VARIABLES=female`
`ch001_ ep005_`
`co007_`
`/ORDER=ANALYSIS.`
- ▶ **Nebo jednodušeji:** `fre`
`female ch001_`
`ep005_ co007_.`

Number of children

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	18349	8,3	9,6	9,6
	1	33275	15,1	17,4	27,0
	2	75949	34,4	39,7	66,7
	3	37883	17,2	19,8	86,6
	4	14752	6,7	7,7	94,3
	5	5731	2,6	3,0	97,3
	6	2576	1,2	1,3	98,6
	7	1258	,6	,7	99,3
	8	666	,3	,3	99,6
	9	273	,1	,1	99,8
	10	210	,1	,1	99,9
	11	78	,0	,0	99,9
	12	63	,0	,0	100,0
	13	28	,0	,0	100,0
	14	15	,0	,0	100,0
	15	9	,0	,0	100,0
	16	9	,0	,0	100,0
17	6	,0	,0	100,0	
Total		191130	86,7	100,0	
Missing	-15. no information	555	,3		
	-13. not asked in this wave	28495	12,9		
	-12. don't know / refusal	397	,2		
	Total	29447	13,3		
Total		220577	100,0		

Filtr důchodkyně s dětmi

- ▶ USE ALL.
- ▶ COMPUTE filter_\$(female = 1 & ch001_ > 0 & ep005_ = 1).
- ▶ VARIABLE LABELS duchodkyne_sdetmi 'female = 1 & ch001_ > 0 & ep005_ = 1 (FILTER) '.
- ▶ VALUE LABELS filter_\$ 0 'Not Selected' 1 'Selected'.
- ▶ FORMATS filter_\$ (f1.0).
- ▶ FILTER BY filter_\$.
- ▶ EXECUTE.

KONTROLA:

Use all.

fre duchodkyne_sdetmi.

Filtr důchodkyně s dětmi

→ Frequencies

Statistics

female = 1 & ch001_ > 0 & ep005_ = 1 (FILTER)

N	Valid	203235
	Missing	17342

female = 1 & ch001_ > 0 & ep005_ = 1 (FILTER)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not Selected	157470	71,4	77,5	77,5
	Selected	45765	20,7	22,5	100,0
	Total	203235	92,1	100,0	
Missing	System	17342	7,9		
Total		220577	100,0		

Asociace mezi příjmem domácnosti a statusem

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	225,059 ^a	3	,000
Likelihood Ratio	225,441	3	,000
Linear-by-Linear Association	38,609	1	,000
N of Valid Cases	187325		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 4772,71.

Symmetric Measures

		Value	Approximate Significance
Nominal	Phi	,035	,000
	Cramer's V	,035	,000
es		187325	

Podmínka χ^2 testu:

- žádná buňka by neměla mít očekávanou četnost <1
- více než 20 % buněk by nemělo mít očekávanou četnost <5 .

Pokud je $< 0,05$ (hladina spolehlivosti α), signifikance, tzn. zkoumané proměnné jsou na sobě závislé.

Síla vztahu

Děkuji za pozornost

Jaroslava.pospisilova@soc.cas.cz

radim.bohacek@cerge-ei.cz