

UK FHS
Historická sociologie, Řízení a supervize
(LS 2011)

Úvod do SPSS (PSPP)

Agregace dat

Jiří Šafr

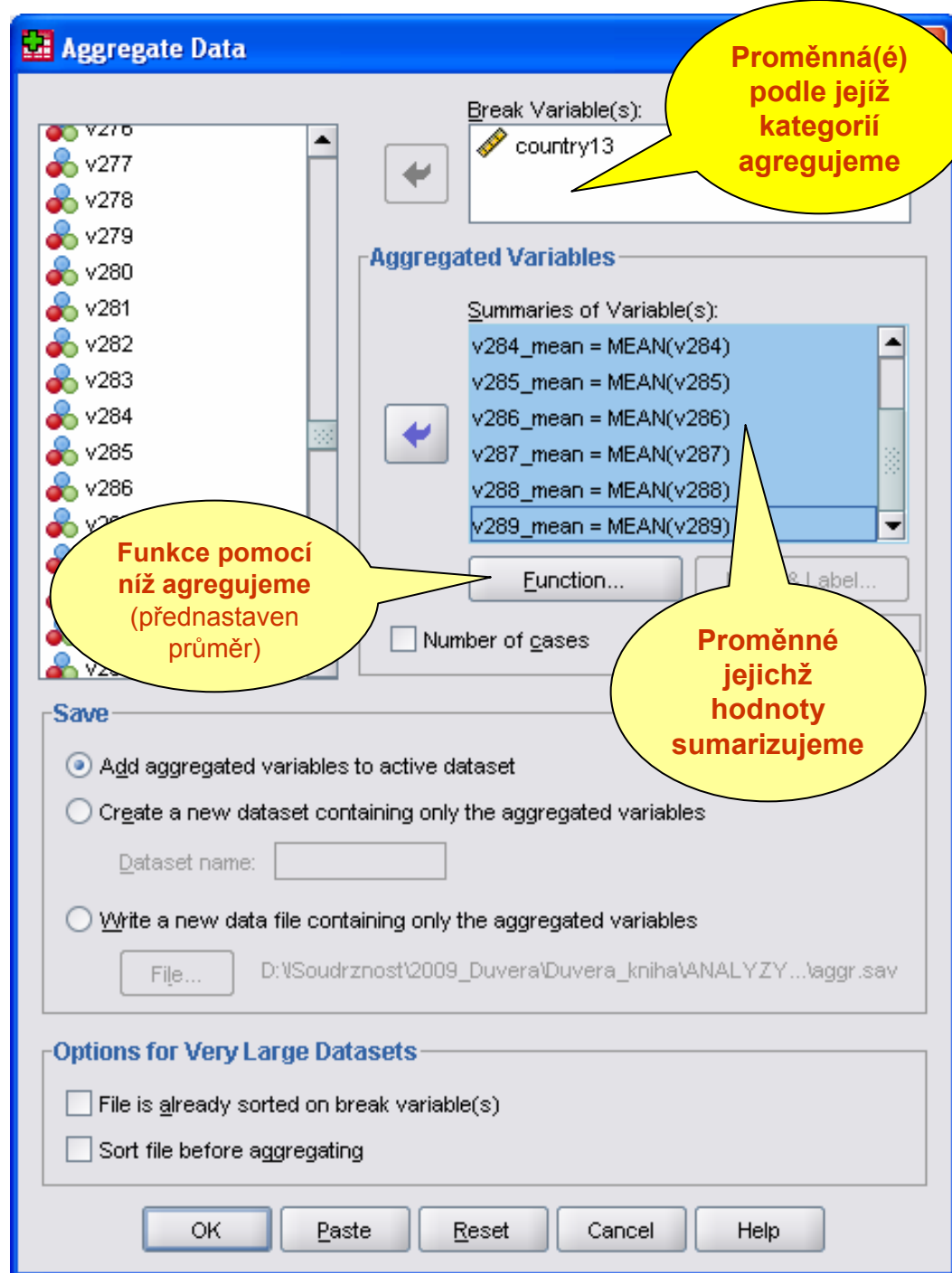
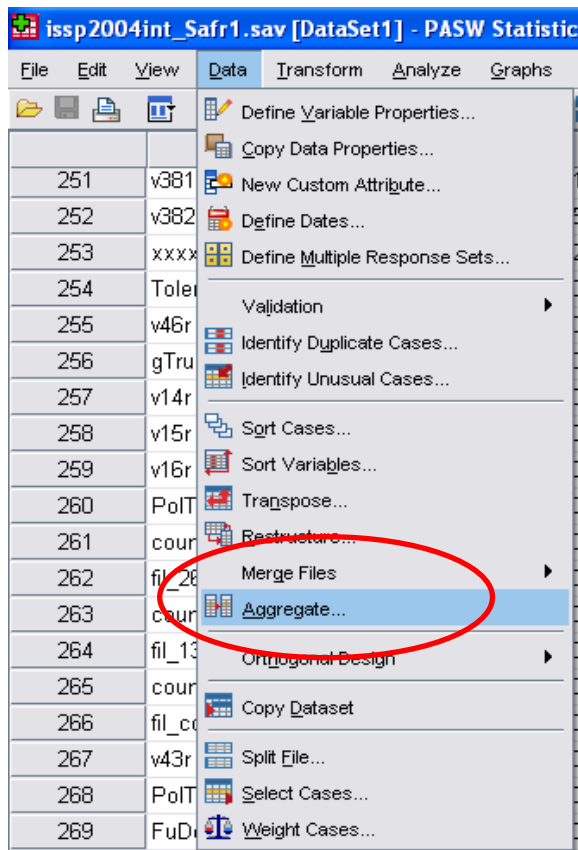
`jiri.safr(AT)seznam.cz`

Poslední aktualizace 6/4/2011

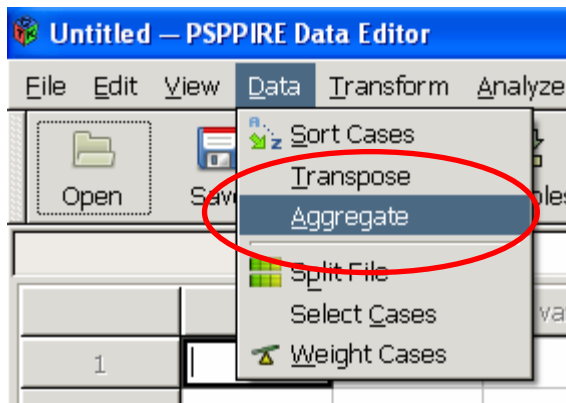
Princip agregace

- Agregování jednotek (originální mikro-data, např. respondenti) do určených skupin. Transformujeme soubor jednotlivců podle hodnot určité proměnné na soubor skupin: jedinci → domácnost, zaměstnanci → organizace, geografické nebo administrativní jednotky, v případě časové řady např. z dní na měsíce.
- **Mikro-data** → **agregovaná data**
x
x → **X**
x
y
y → **Y**
y
- Agregujeme podle určité společné vlastnosti případů, například na základě průměru nebo % určité kategorie (lze také pro směrodatnou odchylku, sumu, nejvyšší/nejnižší hodnotu atd.).

AGGREGATE



v PSPP



Nastavení funkce agregace

Aggregate Data: Aggregate Function

Summary Statistics	Specific Values	Number of cases
<input type="radio"/> Mean	<input type="radio"/> First	<input type="radio"/> Weighted
<input type="radio"/> Median	<input type="radio"/> Last	<input type="radio"/> Weighted missing
<input type="radio"/> Sum	<input type="radio"/> Minimum	<input type="radio"/> Unweighted
<input type="radio"/> Standard Deviation	<input type="radio"/> Maximum	<input type="radio"/> Unweighted missing

Percentages

Above Value:

Below

Inside Low: High:

Outside

Fractions

Above Value:

Below

Inside Low: High:

Outside

Continue Cancel Help

Krom průměru, lze např. procenta v určité kategorii(ích)

Aggregate Data

Break Variable(s):
country13

Aggregated Variables

Summaries of Variable(s):
v284_mean = MEAN(v284)
v285_mean = MEAN(v285)
v286_mean = MEAN(v286)
v287_mean = MEAN(v287)
v288_mean = MEAN(v288)
v289_mean = MEAN(v289)

Function... Name & Label...

Number of cases Name: N_BREAK

Save

Add aggregated variables to active dataset

Create a new dataset containing only the aggregated variables

Dataset name:

Write a new data file containing only the aggregated variables

Aggregate Data: Variable Name ...

PGT(v205 4)

Name:

Label:

Continue Cancel Help

AGGREGATE syntax

***nejprve je třeba data seřadit podle agregující proměnné !.**

`SORT CASES BY v3.`

AGGREGATE

`/OUTFILE='D:\!FHS_HS\2011_LS\AKD2\agr_ISSP04_demokr_zeme.sav'`

`/PRESORTED`

`/BREAK=v3`

`/Tolerance_mean=MEAN(Tolerance)`

`/v46r_mean=MEAN(v46r)`

`/gTrust_mean=MEAN(gTrust)`

`/v14r_mean=MEAN(v14r)`

`/v15r_mean=MEAN(v15r)`

`/v16r_mean=MEAN(v16r)`

`/PolToleran_mean=MEAN(PolToleran)`

`/country15toler_mean=MEAN(country15toler)`

`/fil_country15toler_mean=MEAN(fil_country15toler)`

`/v43r_mean=MEAN(v43r)`

`/PolTrust_mean=MEAN(PolTrust)`

`/FuDem_mean=MEAN(FuDem)`

`/ncActiv_mean=MEAN(ncActiv)`

`/Member_mean=MEAN(Member)`

`/vzd_VS_p '% vysokoškoláků'=PGT(v205 4)`

`/vzd_ZS_p '% základní vzdělání'=PLT(v205 2)`

`/N_BREAK=N.`

Pokud si necháme vytvořit proměnnou **N_BREAK**, tak budeme mít v novém agregovaném souboru absolutní počty v kategoriích proměnné podle níž agregujeme.

Zde počet respondentů ISSP v té které zemi.

Agregovaná data

Příklad ISSP 2004 international:
důvěra, postoje k demokracii
vzdělání podle zemí

Agregovaná data. Příklad: 16 znaků podle zemí

*agr_ISSP04_demokr_zeme.sav [DataSet2] - PASW Statistics Data Editor

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

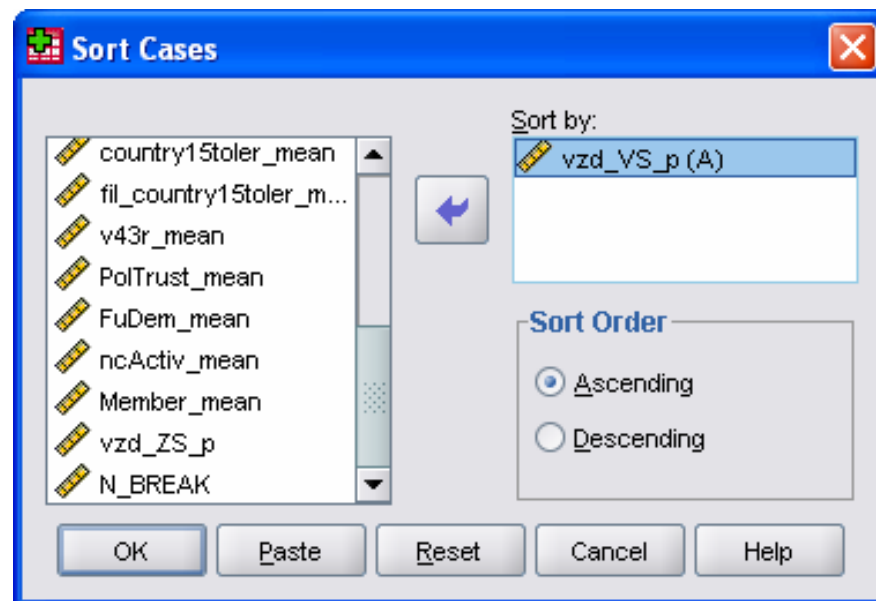
1 : zeme 6

	zeme	zeme\$	Tolerance_mean	v46r_mean	gTrust_mean	v14r_mean	v15r_mean	v16r_mean	PolToleran_mean
1	6	US-United States	.	2,42	63,75	2,95	1,94	2,16	58,95
2	12	NO-Norway	77,18	2,72	70,31	2,02	1,70	2,10	48,82
3	28	FR-France	70,74	2,26	62,63	1,67	1,57	1,64	40,50
4	20	CA-Canada	75,81	2,53	66,37	2,22	1,63	1,71	46,21
5	22	IL-Israel	.	2,25	59,60	2,83	1,81	1,80	53,77
6	41	KR-South Korea	.	2,25	60,89	1,96	1,69	1,88	46,23
7	1	AU-Australia	74,06	2,51	65,04	1,87	1,52	1,65	41,85
8	18	RU-Russia	70,53	2,07	58,62	1,51	1,48	1,42	36,63
9	19	NZ-New Zealand	73,62	2,53	66,15	2,52	1,72	2,00	51,91
10	13	SE-Sweden	.	2,64	66,55	2,08	1,51	1,81	45,31
11	4	GB-Great Britain	.	2,42	63,25	2,08	1,63	1,81	46,00
12	17	BG-Bulgaria	.	1,94	54,76	1,16	1,34	1,24	30,50
13	29	CY-Cyprus	.	2,68	68,86	2,49	1,31	2,01	48,31
14	26	LV-Latvia	73,38	2,01	57,15	1,51	1,54	1,51	37,88
15	37	FI-Finland	.	2,49	64,82	2,47	1,39	1,96	49,22
16	10	IE-Ireland	.	2,43	65,65	2,14	1,61	1,65	44,65
17	32	DK-Denmark	76,64	2,81	71,71	1,98	1,55	1,92	45,57
18	21	PH-Philippines	.	1,87	57,79	2,58	1,63	1,57	48,05
19	24	JP-Japan	.	2,18	62,84	1,40	1,33	1,50	35,09
20	38	MX-Mexico	.	2,40	57,80	2,08	1,72	1,53	44,47
21	16	PL-Poland	.	1,97	53,87	1,93	1,77	1,61	44,17
22	15	SI-Slovenia	.	2,22	58,14	2,04	1,64	1,77	45,38
23	27	SK-Slovak Republ	66,38	2,20	58,12	1,69	1,96	1,49	42,63
24	39	TW-Taiwan	79,79	2,07	62,32	2,03	1,43	1,74	43,11
25	8	HU-Hungary	.	2,39	64,60	1,60	1,38	1,35	35,72

Data View Variable View

Analýzy agregovaných dat

- nejdříve si musíme nový agregovaný soubor otevřít (v SPSS10 nezbytně), automatický název bývá **aggr.sav**, pokud jsme si ho nenechali vytvořit do nového okna s daty.
- Můžeme například:
země seřadit podle
proměnné % VŠ vzdělání
a udělat rychlý výpis dat
→ **SORT CASES**
→ **LIST**



SORT CASES BY vzd_VS_p(D).

LIST zeme\$ vzd_VS_p vzd_ZS_p N_BREAK.

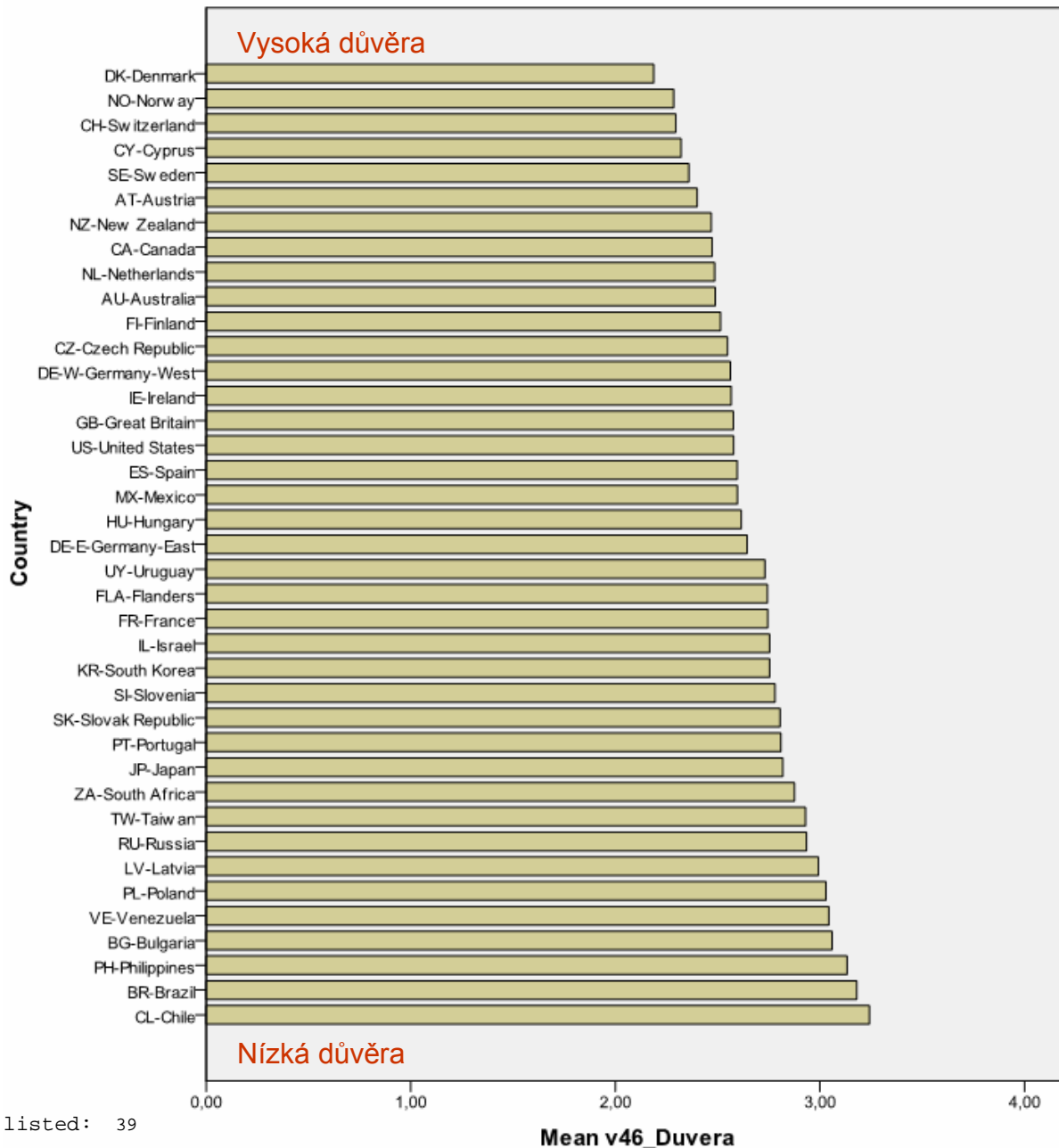
Příkaz LIST pro vybrané proměnné

(vhodný je předtím SORT)

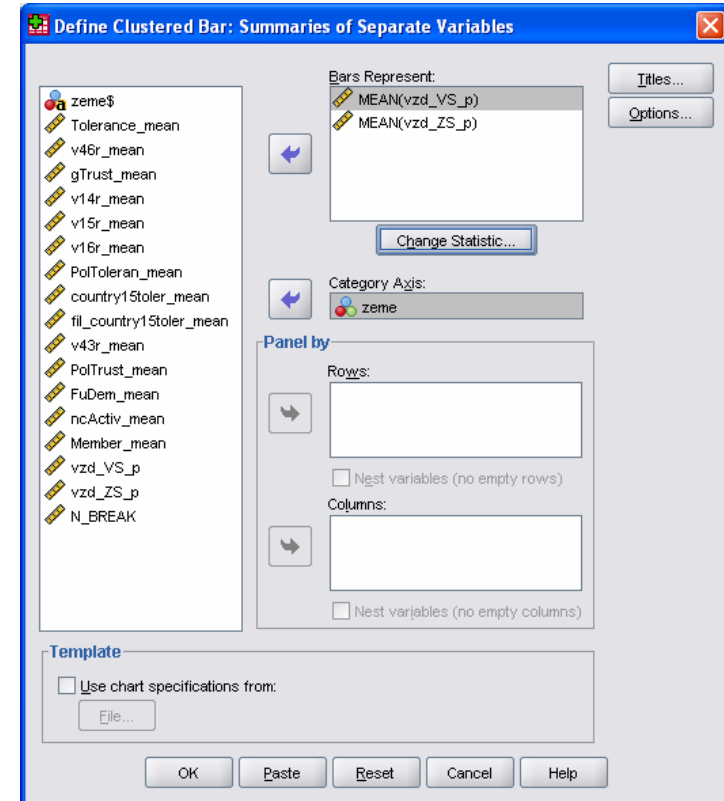
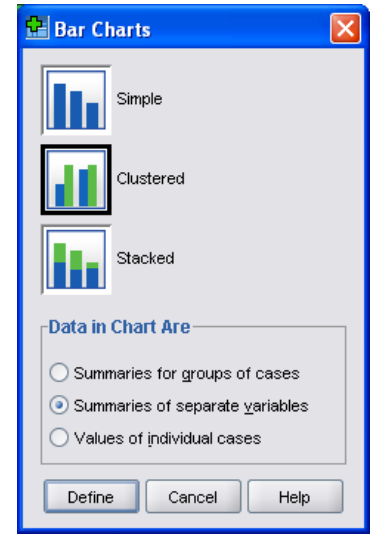
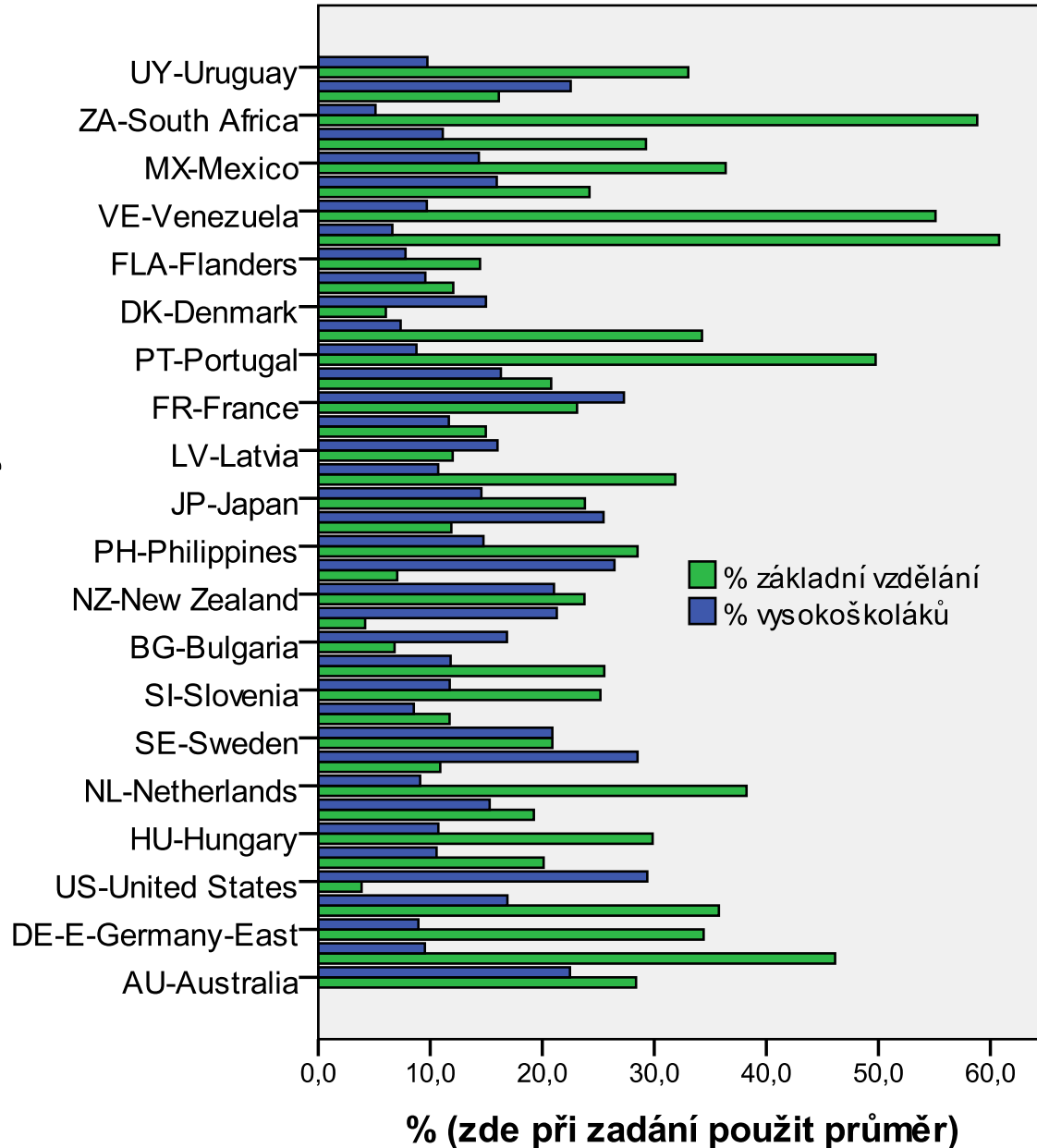
zeme\$	vzd_VS_p	vzd_ZS_p	N_BREAK
US-United States	29,4	3,9	1472
NO-Norway	28,5	10,9	1404
FR-France	27,3	23,1	1419
CA-Canada	26,4	7,1	1211
IL-Israel	25,5	11,9	1184
KR-South Korea	22,5	16,1	1312
AU-Australia	22,5	28,4	1914
RU-Russia	21,3	4,2	1789
NZ-New Zealand	21,0	23,8	1370
SE-Sweden	20,9	20,9	1295
GB-Great Britain	16,9	35,8	853
BG-Bulgaria	16,8	6,8	1121
CY-Cyprus	16,3	20,8	1000
LV-Latvia	16,0	12,0	1000
FI-Finland	15,9	24,2	1354
IE-Ireland	15,3	19,2	1065
DK-Denmark	15,0	6,0	1186
PH-Philippines	14,8	28,5	1200
JP-Japan	14,6	23,8	1343
MX-Mexico	14,3	36,4	1201
PL-Poland	11,8	25,5	1277
SI-Slovenia	11,7	25,2	1054
SK-Slovak Republ	11,7	15,0	1072
TW-Taiwan	11,1	29,3	1781
HU-Hungary	10,7	29,9	1035
ES-Spain	10,7	31,9	2481
AT-Austria	10,6	20,1	1006
UY-Uruguay	9,7	33,0	1108
VE-Venezuela	9,7	55,1	1199
CH-Switzerland	9,6	12,1	1078
DE-W-Germany-Wes	9,5	46,1	896
NL-Netherlands	9,1	38,2	1823
DE-E-Germany-Eas	8,9	34,4	436
PT-Portugal	8,8	49,7	1602
CZ-Czech Republi	8,5	11,7	1322
FLA-Flanders	7,8	14,4	1398
CL-Chile	7,4	34,3	1505
BR-Brazil	6,6	60,8	2000
ZA-South Africa	5,1	58,8	2784

Barchart s průměry 1 znaku

(pro jeho vytvoření ovšem není třeba agregace)

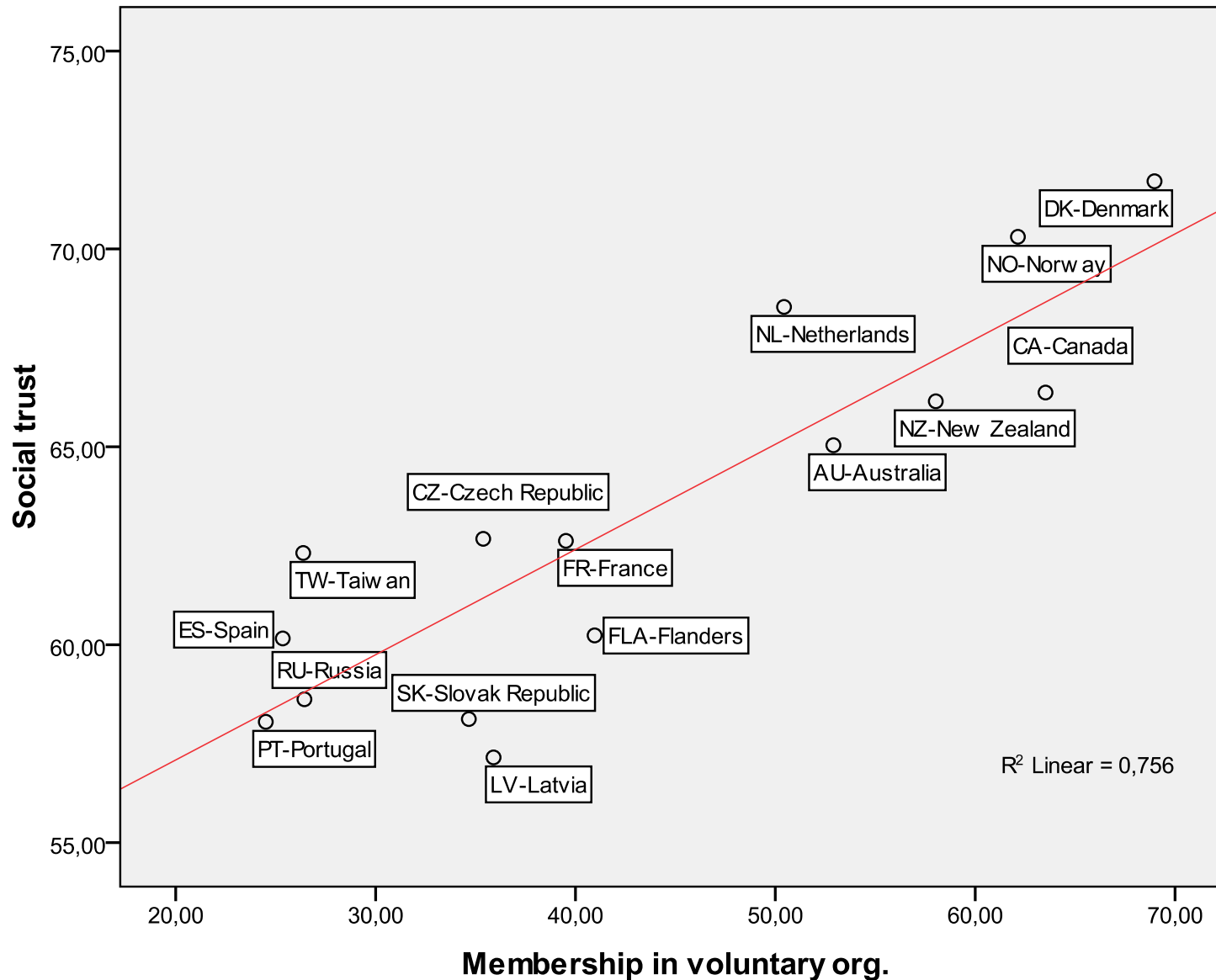


Hodnoty více agregovaných proměnných



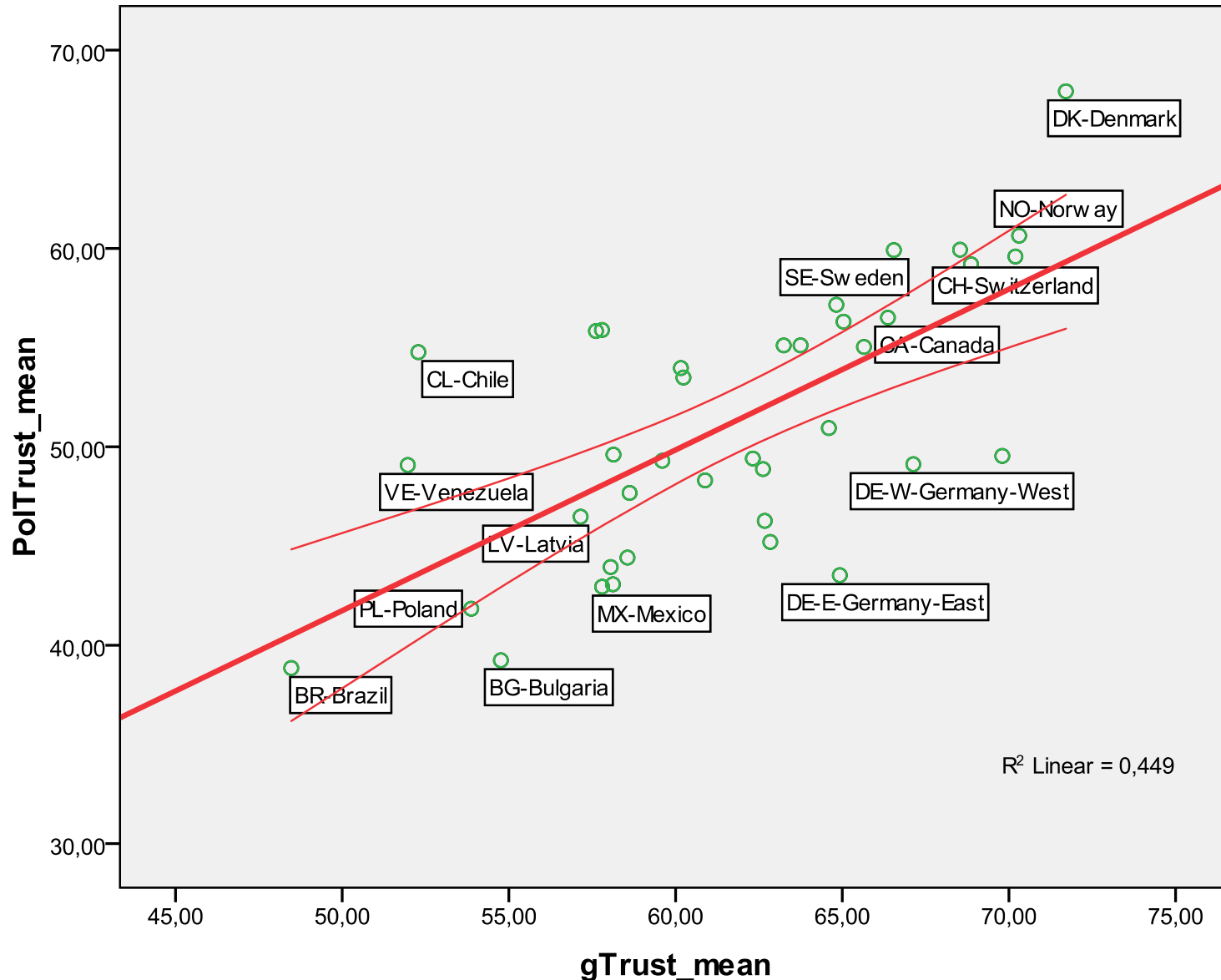
X–Y graf, ScatterPlot s odhadem regresní přímky

Social Trust & Membership in voluntary org.

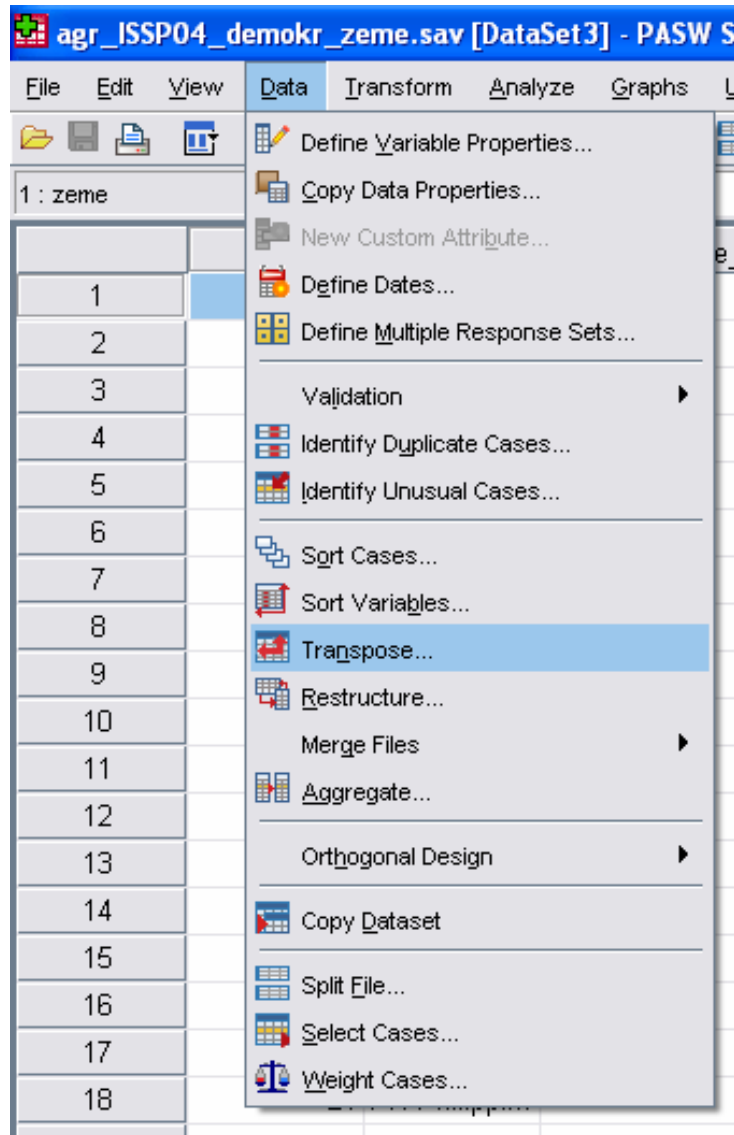


Důvěra v druhé lidi a Důvěra v politické instituce

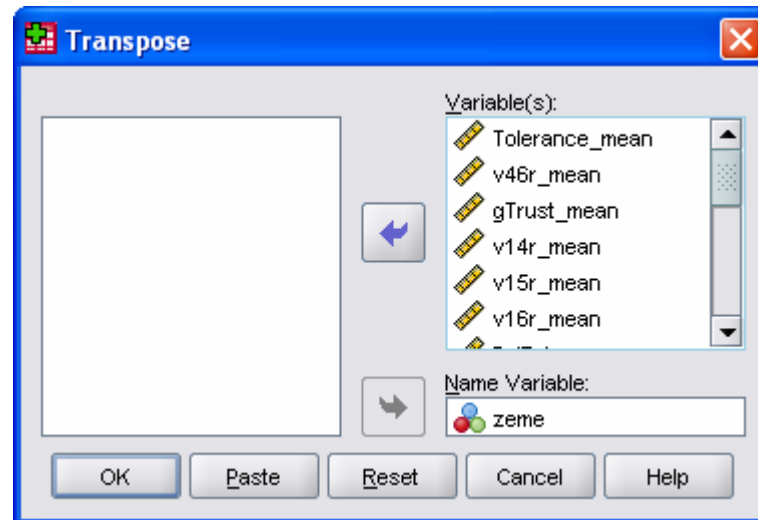
ScatterPlot: s odhadem regresní přímky a pásma 95% CFI



Transponování agregovaného souboru



- zaměnění případů za proměnné (řádků za sloupce) → otočení datové matice o 90st.
- Nutné pro vytvoření některých grafů, v některých analýzách (např. klastrování)



Transponovaná matice dat

Výhodné je použít stringovou proměnnou (zde jména zemí **zeme\$**) k pojmenování nových proměnných.

The screenshot shows the PASW Statistics Data Editor interface. The main window displays a data table with 18 rows and 7 columns. The first row is the header row, and the remaining rows contain numerical data. A red oval highlights the header row and the first row of data. A dialog box titled "Transpose" is open in the bottom right corner, showing a list of variables to be transposed and a field for naming the new variables, which is set to "zeme\$".

	CASE_LBL	US_United_States	NO_Norway	FR_France	CA_Canada	IL_Israel	KR_South_Korea
1	Tolerance_mean	.	77,18	70,74	75,81	.	.
2	v46r_mean	2,42	2,72	2,26	2,53	2,25	2,25
3	gTrust_mean	63,75	70,31	62,63	66,37	59,60	60,89
4	v14r_mean	2,95	2,02	1,67	2,22	2,83	1,96
5	v15r_mean	1,94	1,70	1,57	1,63	1,81	1,69
6	v16r_mean	2,16	2,10	1,64	1,71	1,80	1,88
7	PolToleran_mean	58,95	48,82	40,50	46,21	53,77	46,23
8	country15toler_mean	,00	1,00	1,00	1,00	,00	,00
9	fil_country15toler_mean	,00	1,00	1,00	1,00	,00	,00
10	v43r_mean	2,83	2,87	2,63	2,96	2,70	2,66
11	PolTrust_mean	55,12	60,64	48,88	56,51	49,30	48,31
12	FuDem_mean	6,62	6,68	5,61			
13	ncActiv_mean	33,79	32,55	35,65			
14	Member_mean	55,15	62,14	39,52			
15	vzd_VS_p	29,37	28,50	27,29			
16	vzd_ZS_p	3,87	10,89	23,10			
17	N_BREAK	1472,00	1404,00	1419,00			
18	zeme	6,00	12,00	28,00			

Transpose Dialog Box:

- Variable(s): Tolerance_mean, v46r_mean, gTrust_mean, v14r_mean, v15r_mean, v16r_mean, ...
- Name Variable: zeme\$

Transponování dat → FLIP

```
FLIP VARIABLES=Tolerance_mean v46r_mean gTrust_mean  
v14r_mean v15r_mean v16r_mean PolToleran_mean  
country15toler_mean fil_country15toler_mean v43r_mean  
PolTrust_mean FuDem_mean ncActiv_mean  
Member_mean vzd_VS_p vzd_ZS_p N_BREAK  
/NEWNAMES=zeme.
```

***varianta s využitím stringové proměnné (zde zeme\$)
k pojmenování nových proměnných.**

```
FLIP VARIABLES=Tolerance_mean v46r_mean gTrust_mean  
v14r_mean v15r_mean v16r_mean PolToleran_mean  
country15toler_mean fil_country15toler_mean v43r_mean  
PolTrust_mean FuDem_mean ncActiv_mean Member_mean  
vzd_VS_p vzd_ZS_p N_BREAK zeme  
/NEWNAMES=zeme$.
```

Interpretace agregovaných dat

- Pozor na chybnou interpretaci výsledků týkajících se jedinců (chování, názory) – *mikroanalýza* na základě agregovaných dat – *makroanalýza*!
→ „**ekologicky chybné usuzování**“