

Zpracování dat v kvantitativním šetření

Radoslav Škapa

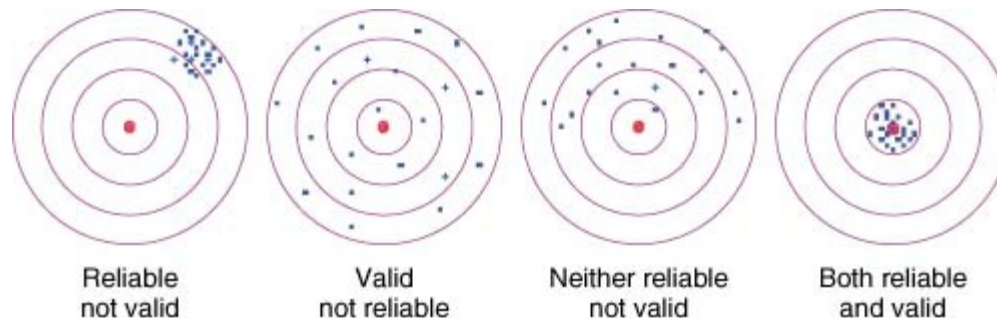


Kvalita výzkumu

- Kvalita výzkumu je vedle použitých metod vyhodnocení, jasné definice cíle, fundované interpretace odvislá zejména kvalitě dat, která má dvě klíčové složky:
 - reprezentativitu (externí validita)
 - a kvalitu měření (interní validita)

Kvalita výzkumu: kvalita měření

- Validita (platnost) – zda test skutečně měří to, co chceme měřit
- Reliabilita (spolehlivost)
 - Reliabilita je předpokladem, aby byl test validní!
 - složky reliability: stabilita v čase, ekvivalenci (různý postup měření – stejný výsledek) a vnitřní konzistenci (Split-Half metoda)



Kvalita výzkumu: kvalita měření

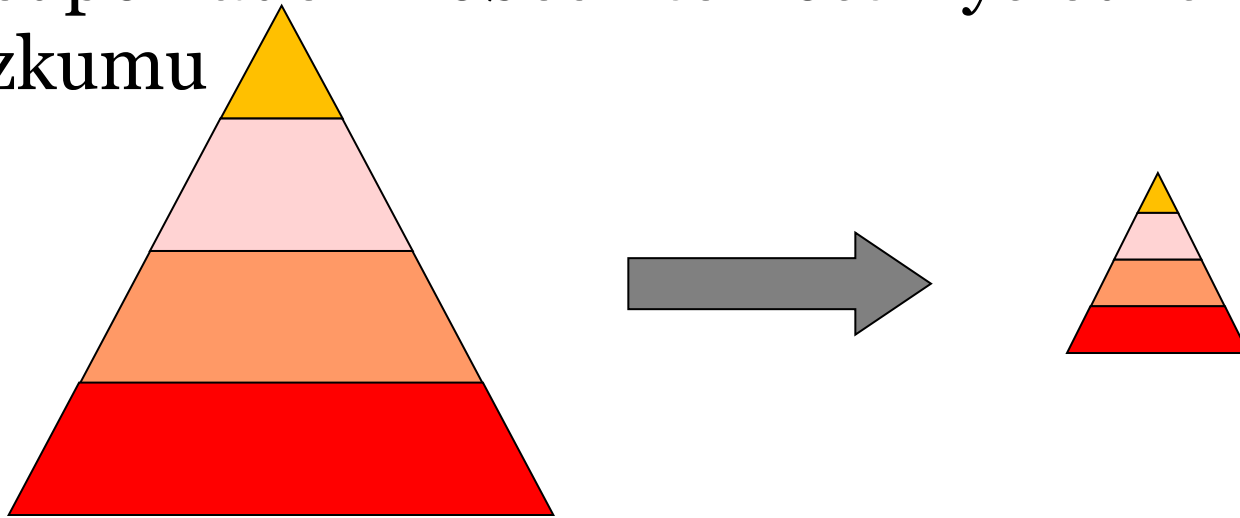
- Měření reliability
 - Měří se snadněji než validita
 - Opakované měření v čase (v krátkém čase se jev nemění)
 - mezi-položková reliabilita (konzistence v odpovědích na baterii otázek)
 - alternativní forma jedné otázky (např. různé pořadí nabízených odpovědí)
 - Více hodnotitelů hodnotí jednu věc – zkoumá se shoda

Kvalita výzkumu: kvalita měření

- Kriteriaální validita – výsledek se porovnává s validizovaným kritériem (současně, retrospektivně). Kde takové kritéria vzít?
 - Příklad: prediktivní modely se porovnaní se skutečností, která nastane (retrospektivní hodnocení).
- Konstruktová validita – zjišťuje zvolený nástroj (ukazatel) to, co mě zajímá?
- Obsahová validita – soulad mezi tím co jsme testovali a tím co jsme testovat měli. Např. Obsahuje zkouškový test otázky na podstatné znalosti z celého učiva? (předpokladem je existence teorie, průzkumu, názory expertů).

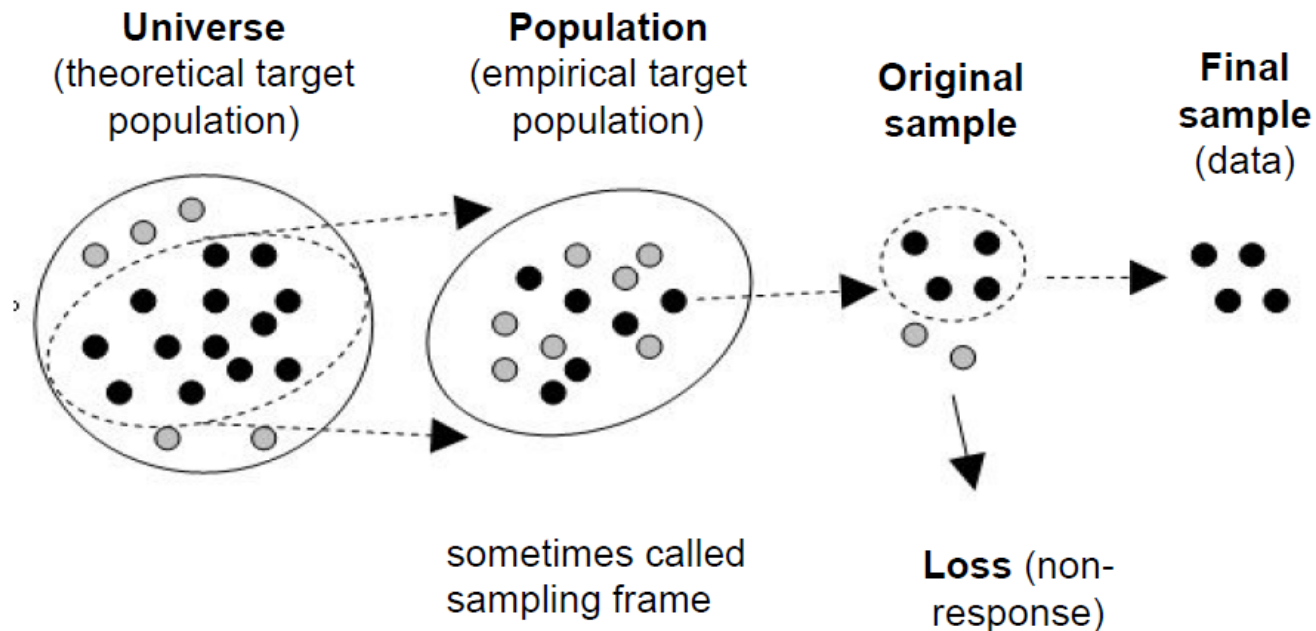
Kvalita výzkumu: reprezentativita

- Míra shody mezi základním a výběrovým souborem z hlediska kvantity i kvality
- Předpokladem zobecnitelnosti výsledků výzkumu



Kvalita výzkumu: výběrový vzorek

Sampling: central concepts



Kvalita výzkumu: výběrový vzorek

- Míra spolehlivosti zobecňování dána kvalitou výběrového vzorku.
- Vždy ale existuje jistá chybovost!
- Větší vzorek vždy lepší než menší (Kdo to zaplatí? Čas?)
- Ideálně zkoumat celý základní soubor – pak netřeba provádět statistické testování výsledků

Kvalita výzkumu: výběrový vzorek

- Míra návratnosti (final sample)?
 - Při písemném dotazování či kontaktování vybraných osob (podniků) běžně 10%
 - Ideálně víc než 50%
 - Nedošlo díky tomu k pokřivení výpovědí? (tzv. non-response bias) – porovnat se znaky základního souboru, které jsou známé + další postupy
- Více: Jindřich Krejčí: Chyba plynoucí z výpadků návratnosti výběrových šetření a statistické dokazování
<http://www.socioweb.cz/index.php?disp=teorie&shw=368&lst=105>

Kvalita výzkumu?

- Jedna z cest jak eliminovat rizika je triangulace
 - triangulace dat – použití více zdrojů dat
 - triangulace výzkumníků – zkušenosti, intersubjektivita
 - triangulace teorií – více způsobů jak data a jevy interpretovat
 - triangulace metod – více metod na zkoumání jednoho jevu

Zdroje chyb



- **Výzkumník:**
 - Chybný či nedostatečný popis základního souboru
 - Chybný výběrový soubor (např. nereprezentativní vzorek, chybná metoda tvorby výběrového souboru).
 - Chybně formulované otázky
- **Tazatel:**
 - Chování vůči respondentům
 - Nedodržení postupu dotazování/podvody
 - Omyly
- **Respondenti:**
 - Neschopnost odpovědět (neznalost, složitá formulace otázek)
 - Neochota odpovědět
 - Neochota odpovědět správně

Příprava dat



Editace a kódování

- struktura datového souboru – značení proměnných, jejich charakter
- i samotné dotazníky je třeba označit, aby byly dohledatelné.

Přepisování dat

- patrně tabulkový procesor (většinou data ve sloupcích)
- kontrola přepisovaných dat – podezřelé hodnoty, či celý dotazník

Kontrola dat

- Jsou hodnoty jednotlivých proměnných smysluplné? Např. extrémní hodnoty, chybějící hodnoty, podezřelé hodnoty (věk 15 let + stav: ženatý)

Typy dat

- **Neparametrické**
 - Nominální (nominal) – např. pohlaví
 - Ordinální (ordinal) – např. preference vyjádřené na škálách, sociální třídy, stupeň vzdělání, toto třídění proměnných z hlediska množství obsažené informace.
- **Metrické (parametrické)**
 - Intervalové (interval) – např. teplota, Likertovy škály – intervaly jsou mezi stupni stejně velké. Nemá ale smysl mluvit o tom, že je např. 2x větší teplota (10 vs. 20 stupňů C). V sociálních výzkumech spíš zřídka.
 - Poměrové (ratio) – např. věk, obrat. Existuje nula. Mnoho statistických testů nerozlišuje mezi intervalovými a poměrovými proměnnými

Analýza dat

Popisná statistika

Jedno- a dvourozměrná analýza

Vícerozměrné analýzy

Interpretace

Třídění 1., 2. a 3. stupně

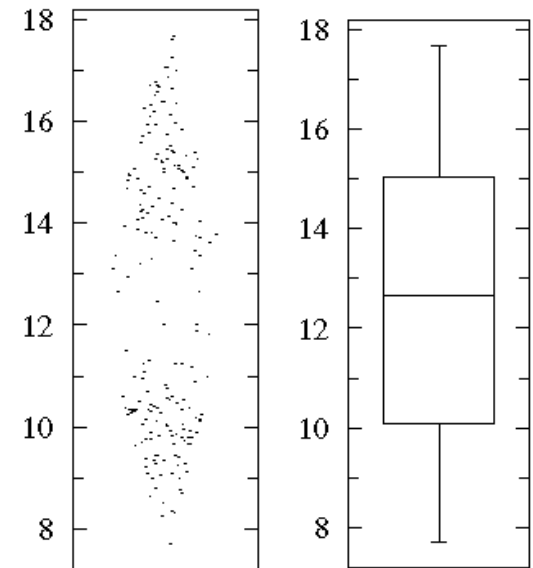
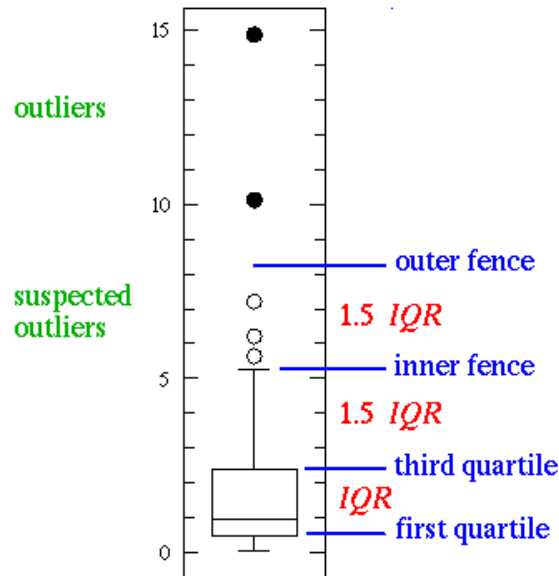
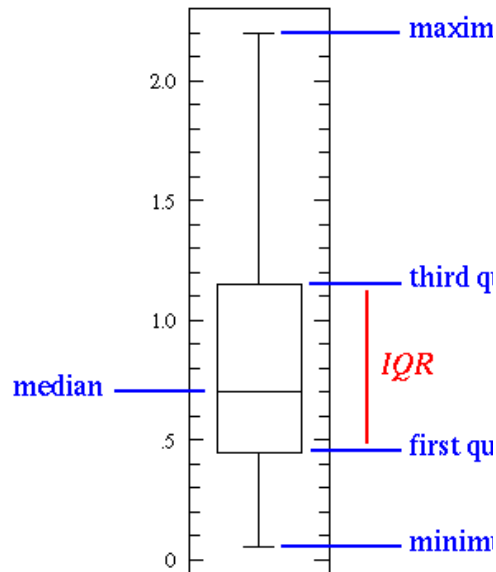


Ukazatelé polohy

Typ proměnné	Přípustné operace
Nominální	Modus
Ordinální	Modus, medián
Intervalové	Modus, medián, průměr
Poměrové	Modus, medián, průměr

- U ordinálních by se neměl počítat průměr. U Likertových škál lze.
- Je vhodné sledovat všechny ukazatele polohy.

Krabicový diagram



Vztahy mezi proměnnými

- Nalezení vztahů je obecným finálním cílem každého výzkumu
- Dvě dimenze vztahu:
 - Velikost (síla) – hodnocení na výzkumníkovi. Obecně ve společenských vědách se za silné vazby považují už nižší koeficienty asociace (např. 0,7) než přírodní vědy. Příklady Pearsonův produktový koeficient korelace.
 - Spolehlivost (reliabilita, pravdivost) – pravděpodobnost, že výsledek není náhodný. Spolehlivost s jakou lze výsledek zobecnit na základní soubor. Měří se pomocí „p-value“ (statistical significance) – pravděpodobnosti chyby. Např. p-value=0,05 znamená 95% spolehlivost.

Vztahy mezi proměnnými

- Z jiného pohledu: $p\text{-value}=0,05$ znamená např. že cca při 20 měřeních korelací nesouvisejících proměnných nám jedna vyjde spolehlivá. (tzv. chyba 1. typu).
- Existuje vztah mezi silou a spolehlivostí vypočteného vztahu (příklad. porodnice)
- Ve stejně velkém vzorku, silnější vztahy víc spolehlivé.
- K prokázání slabých vztahů je třeba velké vzorky. (K prokázání neexistence žádného vztahu – prozkoumat téměř celou populaci). (příklad – slabě vychýlená mince).
 - \Rightarrow ve velkých vzorcích i slabé vztahy budou statisticky významné – proto při interpretaci se vždy zamyslet, zda je takový vztah dostatečně silný, aby mělo smysl o něm mluvit.



Jak se počítá spolehlivost?

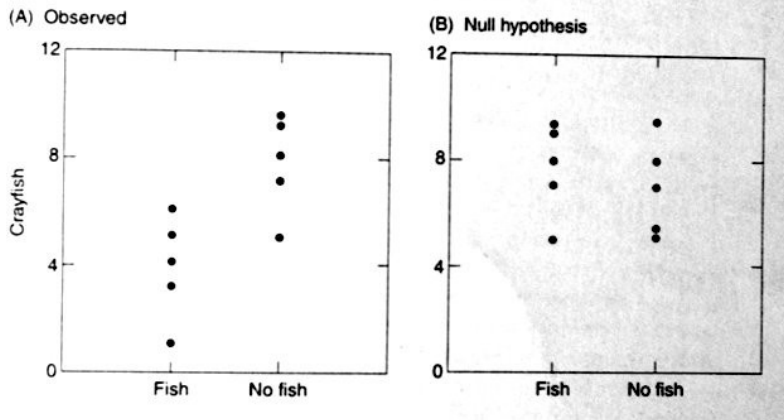


Table 5.1

Stream	Fish	Crayfish	FISH 1	FISH 2	FISH 3
1	+	1	+	+	+
2	-	5	-	+	-
3	+	3	+	+	+
4	-	7	-	-	-
5	+	4	+	-	-
6	-	8	+	-	+
7	+	5	-	-	-
8	-	9	+	-	+
9	+	6	-	+	+
10	-	9.5	-	+	-

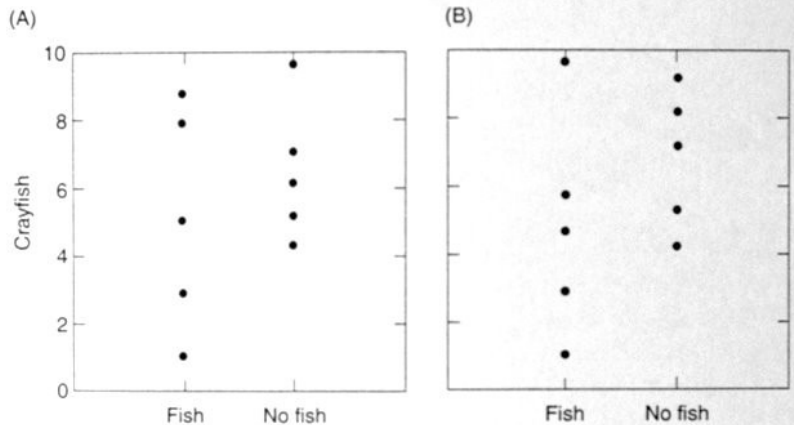
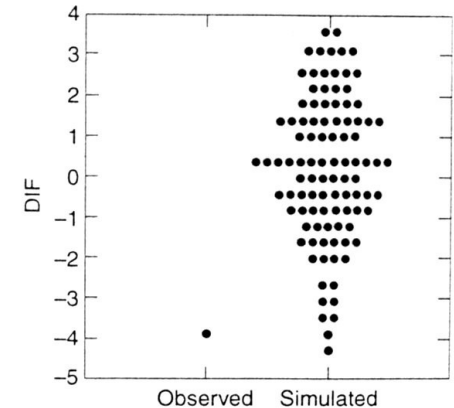


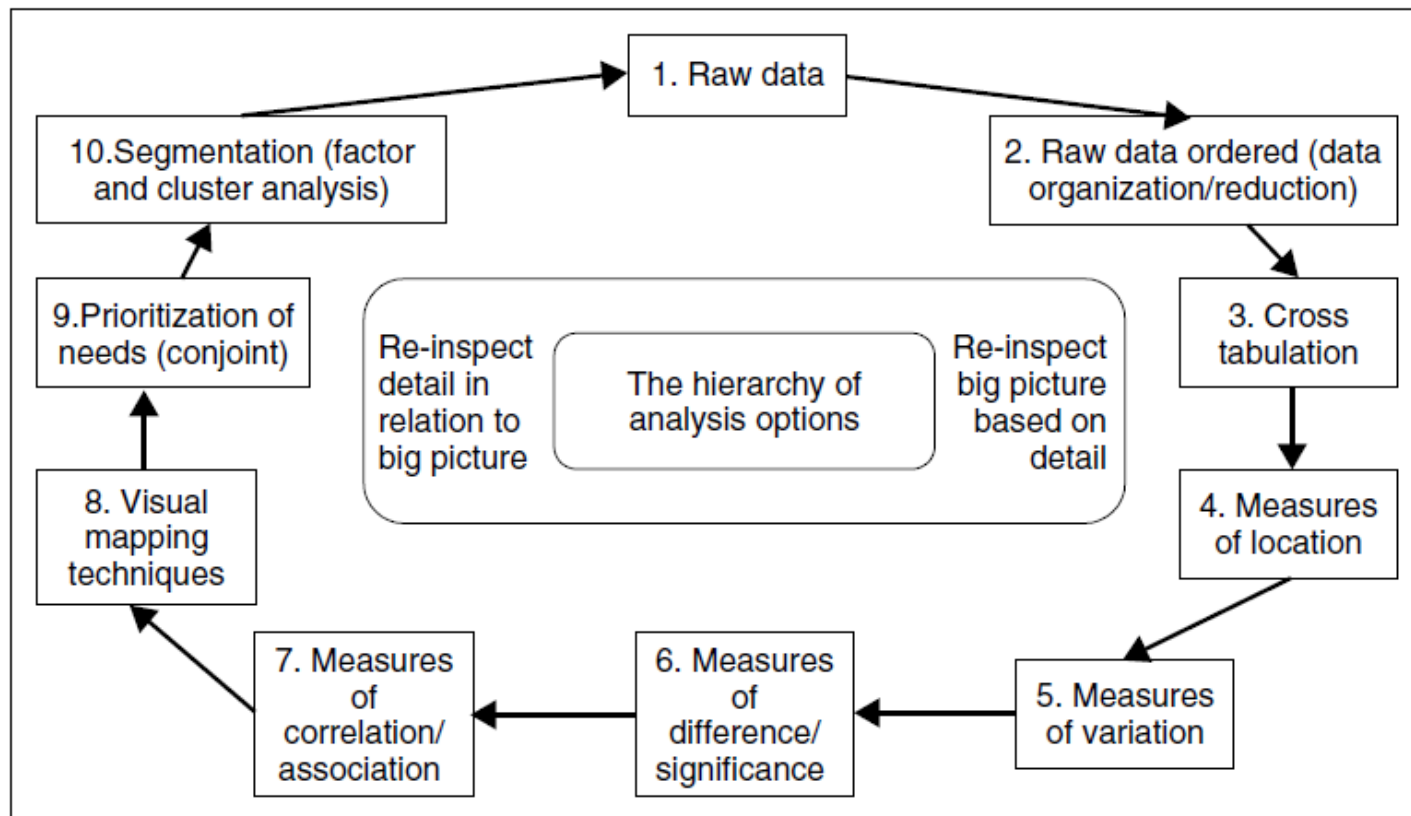
Table 5.2

Graph	DIF	Graph	DIF
1	2.70	11	2.1
2	1.90	12	2.7
3	-0.09	13	-3.1
4	2.50	14	1.5
5	-2.70	15	-1.1
6	1.70	16	-0.1
7	0.50	17	1.3
8	1.10	18	-0.5
9	1.50	19	-0.7
10	3.10	20	-0.3

Figure 5.3



Postup statistického vyhodnocování



Které analýzy v práci použít?

- Třídění 1. a 2. stupně + alespoň několik analýz 3. stupně:
 - Popisná statistika: (analýza četností, polohy, variability)
 - Kontingenční tabulky
 - Rozdíly ve středních hodnotách (t-test, Mann-Whitney test – ordinální data)
 - Korelace (Pearson, Spearman (Kendall) – ordinální data)
- Nezapomenout na interpretaci výsledků
- Ideálně další a náročnější metody – vícerozměrná regrese, shluková analýza, diskriminační analýza, conjoint analyza, faktorová analýza. (tyto je třeba samostatně nastudovat, použít vhodně vzhledem k cíli a sestavit dotazník způsobem, abyste metodu mohli využít)



Software

- MS Excel – doplněk Analýza dat
 - **XLStatistics**
<http://www.deakin.edu.au/~rodneyc/XLStatistics/>
- Statistica – licence MU
- SPSS – Multilicence MU
- Statgraphics – zaměřený spíš na průmysl. Výhodou jsou automatické komentáře k výsledkům.

Literatura

***I*ASTAT - INTERAKTIVNÍ UČEBNICE STATISTIKY**

<http://iastat.vse.cz/>

Štatistický navigátor

<http://rimarcik.com/navigator/>

StatSoft, Inc. (2010). Electronic Statistics Textbook.

<http://www.statsoft.com/textbook/>

Miroslav Disman: Jak se vyrábí sociologická znalost

http://www.jakubholy.net/humanities/disman-soc_znalost.html

Martin Kreidl: Metody měření reliability a validity.

<http://www.socioweb.cz/index.php?disp=teorie&shw=153&lst=106>

**Petr Mareš, Ladislav Rabušic: Studijní materiály pro předět
SOC708**

<https://is.muni.cz/auth/el/1423/podzim2005/SOC708/um/?info=1>