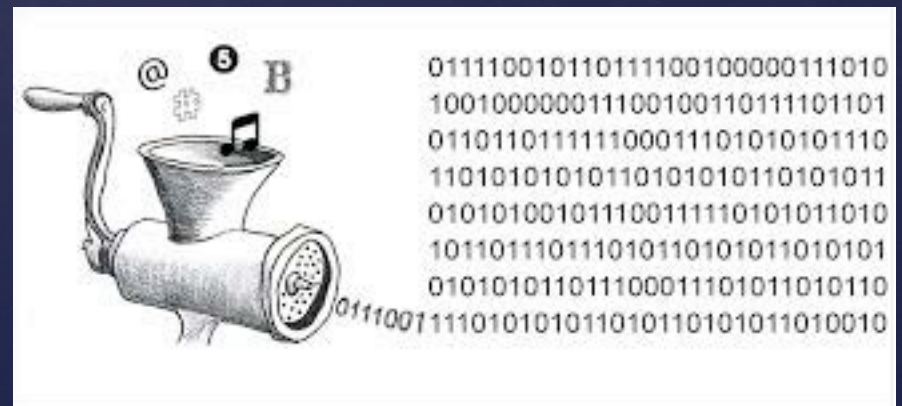


Štandardizácia údajov

{ Prof. MUDr. Martin Rusnák, CSc
Prof. MUDr. Martin Rusnák, CSc
PhDr. Marek Psota, PhD



- ⌘ Zopakovať si základy demografie z prvého ročníka a prečítať si kapitolu 7 *Štúdium reprodukcie populácií ľudí*
- ⌘ Niečo z histórie
- ⌘ Skreslenie a vychýlenie
- ⌘ Metódy štandardizácie
- ⌘ Očakávaná dĺžka života – opakovanie z demografie a samoštúdium
- ⌘ Súhrn

Obsah



- ⌘ Často je potrebné porovnať úmrtnosti v dvoch či viacerých populáciách, povedzme dvoch krajinách, alebo okresoch či mestách. Vieme však, že v jednej z nich žije podstatne viac detí a mladých ľudí ako v druhej, kde je zas málo detí a veľa starých ľudí. Je logické, že tam, kde je viac starých ľudí, bude úmrtnosť vyššia.
- ⌘ Hrubá úmrtnosť odráža rovnako rozdiely spôsobené samotnou úmrtnosťou, ako aj rozdiely vo vekovom zložení obyvateľstva.
- ⌘ Spôsob, ako riešiť tento problém, sa nazýva štandardizácia a často sa používa pri porovnávaní ukazovateľov zdravia. Porovnávanie úmrtností bolo základnou motiváciou zavedenia štandardizácie. Hoci je možné štandardizovať vzhľadom na rôzne premenné, ktoré môžu vplývať na jej vývoj, vek sa najčastejšie používa na štandardizáciu úmrtnosti pre jeho vplyv na chorobnosť a úmrtnosť.



Načo je štandardizácia



- ⌘ V polovici 19. storočia si odborníci na zdravie verejnosti v Anglicku začali uvedomovať, že hrubé miery sú nevhodné na porovnávanie ukazovateľov zdravia obyvateľov v skupinách, kde vekové rozdelenie bolo výrazne odlišné.
- ⌘ Diskusia sa rozdielmi vo veku. V prednáške pre Štatistickú spoločnosť v Londýne pán **Edwin Chadwick**, ktorého poznáme ako jedeného z prvých reformátorov zdravia verejnosti v Anglicku, navrhol použitie ukazovateľa „stredný vek úmrtia“ ako nástroj porovnania zdravotného stavu rôznych častí Londýna. Tento index podľa neho predstavoval skutočný pohľad na vekovo špecifické riziko umrieť.

Niečo z histórie

& Neison, ktorý sa zaoberal poistnou matematikou, nesúhlasil s jeho výpočtami a dokázal jeho omyl.

& Uviedol aj riešenie, ktoré dodnes označujeme tak, ako to on navrhol, teda priama a nepriama štandardizácia, rovnako ako termín štandardná populácia.

& Už v roku 1883 v správe o populácii v Anglicku bola použitá metóda priamej štandardizácie podľa Neisona a ako štandard boli použité výsledky sčítania ľudu v Anglicku a Walese z roku 1881.

& SZO v roku 2001 zaviedla medzinárodný štandard, ktorý používame dodnes.

MIHL. Legal Treatises, 1800-1926
Statistics of crime in
England and Wales : for
the years 1842, 1843,
and 1844.
F. G. P. Neison



Oprava



- ⌘ Niekedy pozorujeme príčinný vzťah medzi premennými, ktorý v skutočnosti nie je prítomný.
- ⌘ Premenné, ktoré skresľujú vzťahy medzi ďalšími premennými, ktoré sú predmetom nášho skúmania, sa nazývajú **mätúce premenné (confounders)**.
- ⌘ Mätúca premenná musí byť vo vzťahu k ochoreniu alebo k stavu, o ktorý sa zaujímame a tiež k faktoru rizika, ktorý študujeme.
- ⌘ Pokiaľ sa vychýlenie neodstráni, sú výsledky vychýlené.

Skreslenie a vychýlenie



- ⌘ V štatistike považujeme proces za vychýlený, ak nie sú všetky možné výstupy rovnako pravdepodobné; v takom prípade hovoríme, že máme výchylku od najpravdepodobnejšieho výstupu.
- ⌘ Keď hádžeme kockou, automaticky predpokladáme, že pri veľkom počte hodov je rovnaká pravdepodobnosť padnutia každej strany kocky.
- ⌘ Hod kockou je nespravodlivý, ak nie sú všetky dosiahnuté skóre rovnako možné. Ak po mnohých hodoch kockou zaznamenáme viac trojok alebo pätiok a veľmi málo iných výsledkov, môžeme predpokladať, že kocka bola vychýlená - biased - alebo sfaľšovaná.

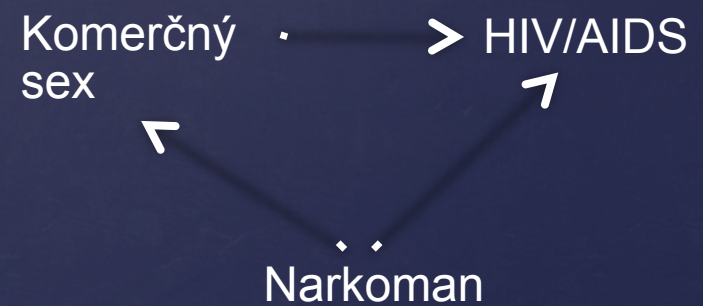
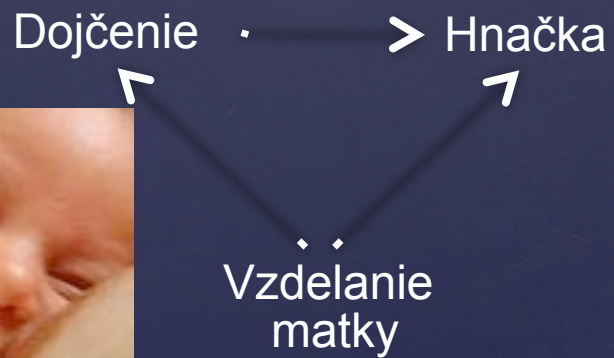
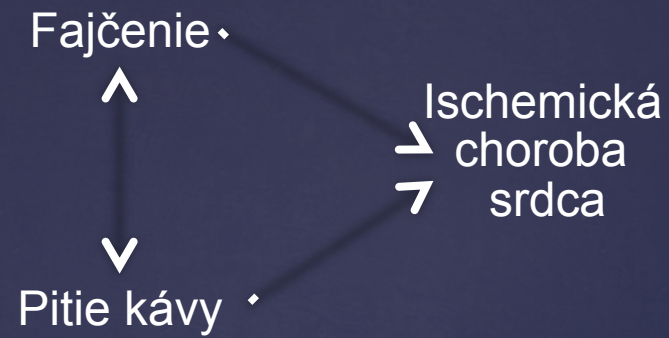
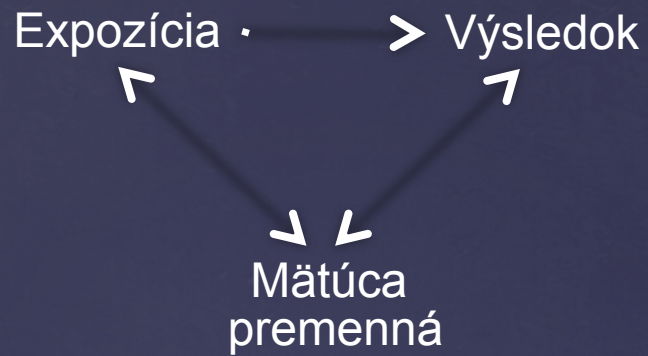
Bias, Vychýlenie, Skreslenie

- ⌘ Systematická (jednostranná) odlišnosť merania od skutočnej hodnoty (synonymum systematická chyba).
- ⌘ Variácie štatistických súhrnných mier (priemery, proporcie, miery asociácie atď.), od ich skutočných hodnôt v dôsledku systematickej odchýlky meraní, iné slabiny v zbere dát, alebo nedostatky v návrhu štúdie alebo analýzy.
- ⌘ Odchýlka záverov od skutočnosti v dôsledku nedostatkov v návrhu štúdie, zbere údajov alebo v analýze alebo interpretácii výsledkov.
- ⌘ Tendencia postupov (návrh štúdie, zber dát, analýza, interpretácia, hodnotenie, alebo publikácia) prinášať výsledky alebo závery, ktoré sa odchyľujú od skutočnosti.
- ⌘ Zaujatosť vedúca k vedomému alebo nevedomému výberu postupov štúdia, ktoré sa líšia od skutočnosti v určitom smere alebo smerujú k jednostrannosti pri interpretácii výsledkov.

Príčiny skreslenia

- ⌘ Stav, keď nie sú účinky dvoch procesov oddelené. Zmiešanie zdanlivého vplyvu expozície na riziká, ktoré prináša spojenie s inými faktormi, ktoré môžu ovplyvniť výsle-dok.
- ⌘ Vzťah medzi účinkami dvoch alebo viacerých príčinných faktorov, ktoré sa pozo-rovali v súbore dát takým spôsobom, že nieje logicky možné oddeliť príspevok k účinku každého z jednotlivých kauzálnych faktorov.
- ⌘ Situácia, keď je miera vplyvu expozície na riziko skreslená vzťahom expozície na iný faktor (y), ktoré ovplyvňujú skúmaný výsledok.

Confounding - zmätenie



& Priama
& Nepriama

Metódy štandardizácie

Age groups	Population I	No. of deaths in each age group	Age specific death rate		Age groups	Population II	No. of deaths in each age group	Age specific death rate

Postup pri priamej štandardizácii

NAING, N. N. 2000. Easy way to learn standardization : direct and indirect methods. *Malays J Med Sci*, 7, 10-5.

	Bratislava I (a)			Bratislava V (b)			B V/BI
Veková skupina (i)	Počet úmrtí x_{ia}	Úmrtnosť /1000 p_{ia}	Populácia n_{ia}	Počet úmrtí x_{ib}	Úmrtnosť /1000 p_{ib}	Populácia n_{ib}	Populácia n_{ib}/n_{ia}
0--14	6	1.3	4 636	5	0.4	12 617	2.7
15--59	63	2.7	23 595	241	2.9	83 491	3.5
60+	490	38.9	12 597	283	13.2	21 464	1.7
Spolu	559	13.7	40 828	529	4.5	117 572	2.9

Počty zomretých a počet obyvateľov v Bratislava I a Bratislava IV v roku 2009, ženy a muži spolu. Zdroj: ŠÚ SR

Vek (i)	Populácia BA I + BA V $P_i = n_{iI} + n_{iB}$	Bratislava I			Bratislava V		
		Úmrtnosť /1000 p_{iI}	Očakáva- ný počet úmrtí e_{iI}	Štandar- dizovaná úmrtnosť sdr_{iI}	Úmrtnosť /1000 p_{iB}	Očakáva- ný počet úmrtí e_{iB}	Štandar- dizovaná úmrtnosť sdr_{iB}
0-14	17 253	1.3	22.3		0.4	6.8	
15-59	107 086	2.7	285.9		2.9	309.1	
60+	34 061	38.9	1 324.9		13.2	449.1	
Spolu	158 400		1 633.2	0.01031	4.499	712.7	0.00450

Výpočet štandardizovanej úmrtnosti
metódou priamej štandardizácie

	A	B	C	D	E	F	G	H	
1		Bratislava I (a)			Bratislava V (b)			B V/BI	
2	Veková skupina (i)	Počet úmrtí X_{is}	Úmrtnosť /1000 P_{is}	Populácia N_{is}	Počet úmrtí X_{ib}	Úmrtnosť /1000 P_{ib}	Populácia N_{ib}	Populácia N_{ib}/N_{is}	
3	0-14	6.0	1.3	4 636.0	5.0	0.4	12 617.0	2.7	
4	15-59	63.0	2.7	23 595	241.0	2.9	83 491.0	3.5	
5	60+	490.0	38.9	12 597	283.0	13.2	21 464.0	1.7	
6	Spolu	559.0	13.7	40 828	529.0	4.5	117 572	2.9	
7									

$$=B3/D3*1000$$

$$=E3/G3*1000$$

$$=G3/D3$$

$$=B4/D4*1000$$

$$=E4/G4*1000$$

$$=G4/D4$$

$$=B5/D5*1000$$

$$=E5/G5*1000$$

$$=G5/D5$$

$$=SUM(B3:B5)/D6*1000$$

$$=SUM(E3:E5)/G6*1000$$

$$=G6/D6$$

Pozn. Súčtom obyvateľov z oboch oblastí podľa vekových skupín sme získali spoločnú, alebo štandardnú populáciu P_i , na ktorú prepočítame počty úmrtí z jednotlivých oblastí.

	A	B	C	D	E	F	G	H
1		Bratislava I				Bratislava V		
	Veková skupina (i)	Populácia BA I + BA IV $P = N_{ib} + N_{iv}$	Úmrtnosť /1000 P_{ib}	Očakávaný počet úmrtí E_{ib}	Štandardizovaná úmrtnosť sdr_{ib}	Úmrtnosť /1000 P_{ib}	Očakávaný počet úmrtí E_{ib}	Štandardizovaná úmrtnosť sdr_{ib}
2								
3	0-14	17 253.0	1.3	22.4	0.0013	0.4	6.9	0.0004
4	15-59	107 086.0	2.7	289.1	0.0027	2.9	310.5	0.0029
5	60+	34 061.0	38.9	1 325.0	0.0389	13.2	449.6	0.0132
6	Spolu	158 400.0	13.7	2 170.1	0.0137	4.5	712.8	0.0045
7								

sdr je v
Starom meste
(Bratislava I)
výrazne
vyššia ako v
Petržalke
(Bratislava V).

$$=C3*B3/1000$$

$$=D3/B3$$

$$=F3*B3/1000$$

$$=G3/B3$$

$$=C4*B4/1000$$

$$=D4/B4$$

$$=F4*B4/1000$$

$$=G4/B4$$

$$=C5*B5/1000$$

$$=D5/B5$$

$$=F5*B5/1000$$

$$=G5/B5$$

$$=C6*B6/1000$$

$$=D6/B6$$

$$=F6*B6/1000$$

$$=G6/B6$$

Vychádzajúc z koncepcie relatívneho rizika, môžeme tieto úmrtnosti interpretovať ako pravdepodobnosť umrieť a ich pomer ako relatívne riziko úmrtia:

$$RR = \frac{sdr_a}{sdr_b} = \frac{0,01}{0,0045} = 2,29$$

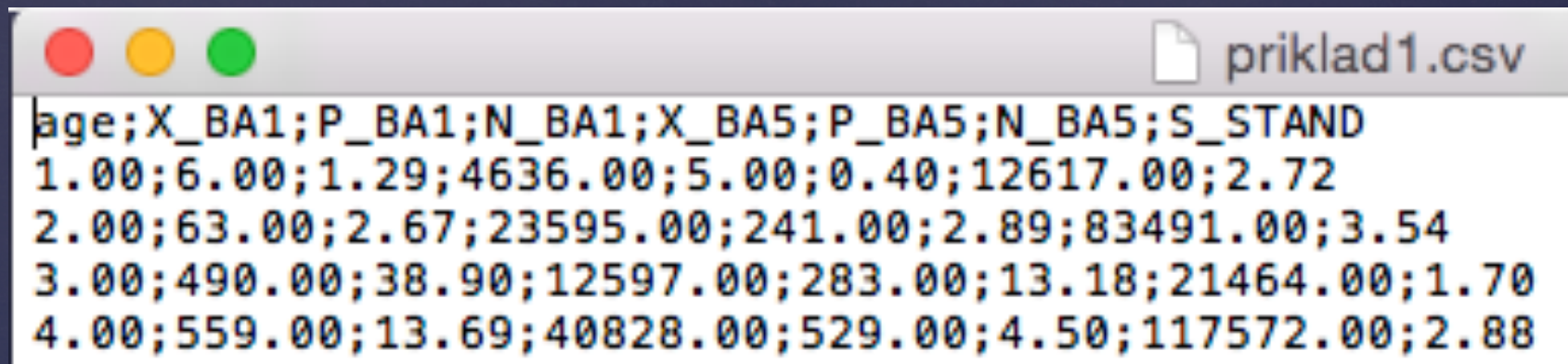
Interpretácia

Epitools

Údaje pripravíme v Excel, nadpíšeme názvy premenných a odstránime medzery

age	X_BA1	P_BA1	N_BA1	X_BA5	P_BA5	N_BA5	S_STAND
1.00	6.00	1.29	4636.00	5.00	0.40	12617.00	2.72
2.00	63.00	2.67	23595.00	241.00	2.89	83491.00	3.54
3.00	490.00	38.90	12597.00	283.00	13.18	21464.00	1.70
4.00	559.00	13.69	40828.00	529.00	4.50	117572.00	2.88

Zapamätáme ako .csv



```
age;X_BA1;P_BA1;N_BA1;X_BA5;P_BA5;N_BA5;S_STAND
1.00;6.00;1.29;4636.00;5.00;0.40;12617.00;2.72
2.00;63.00;2.67;23595.00;241.00;2.89;83491.00;3.54
3.00;490.00;38.90;12597.00;283.00;13.18;21464.00;1.70
4.00;559.00;13.69;40828.00;529.00;4.50;117572.00;2.88
```

Príprava údajov

```
priama<- read.csv(file="priklad1.csv", header=TRUE, sep=";", dec=".")
```

```
> priama
```

```
  age X_BA1 P_BA1 N_BA1 X_BA5 P_BA5 N_BA5 S_STAND
1   1    6 1.29 4636    5 0.40 12617    2.72
2   2   63 2.67 23595  241 2.89 83491    3.54
3   3  490 38.90 12597  283 13.18 21464    1.70
4  559 13.69 40828  529  4.50 117572    2.88
```

Štandardná populácia

```
> popstdn <- priama$N_BA1[1:3]+priama$N_BA5[1:3]
```

```
> popstdn
```

```
[1] 17253 107086 34061
```

Prečítanie do {R}

⌘ **ageadjust.direct(count, pop, rate = NULL, stdpop, conf.level = 0.95)**

⌘ **count** vektor udalostí, teda úmrtí

⌘ **pop** vektor populácie

⌘ **rate** úmrtnosti

⌘ **stdpop** štandardná populácia

⌘ **conf.level** úroveň istoty (nastavené = 0.95)

⌘ `> direct_BA1 <- ageadjust.direct(priama$X_BA1[1:3], priama
$N_BA1[1:3], rate=NULL, popstdn)`

⌘ `direct_BA5 <- ageadjust.direct(priama$X_BA5[1:3], priama
$N_BA5[1:3], rate=NULL, popstdn)`

Volanie funkcie z EPITOOLS

```
& > direct_BA1
```

```
crude.rate  adj.rate      lci      uci  
0.013691584 0.010310384 0.009456982 0.011230608
```

```
> direct_BA5
```

```
crude.rate  adj.rate      lci      uci  
0.004499371 0.004829766 0.004424762 0.005262736
```

	Hrubá miera úmrtnosti	Štandardi zovaná úmrtnosť	Dolný interval istoty	Horný interval istoty
BA 1	0.014	0.01	0.009	0.011
BA 5	0.004	0.005	0.004	0.005

RR = 0.01/0.005

RR = 2

Výsledok

SZO štandardná populácia

Veková skupina	Európska populácia	SZO populácia
0-4	8.00	8.86
5-9	7.00	8.69
10-14	7.00	8.60
15-19	7.00	8.47
20-24	7.00	8.22
25-29	7.00	7.93
30-34	7.00	7.61
35-39	7.00	7.15
40-44	7.00	6.59
45-49	7.00	6.04
50-54	7.00	5.37
55-59	6.00	4.55
60-64	5.00	3.72
65-69	4.00	2.96
70-74	3.00	2.21
75-79	2.00	1.52
80-84	1.00	0.91
85+	1.00	0.63
Total	100.00	100.00

- ⌘ Nepriama štandardizácia sa zvolí vtedy, keď nie je dostatok údajov pre priamu, používa sa menej často.
- ⌘ Vychádza zo štandardnej množiny vekovo špecifických úmrtností, ktoré sa vzťahujú na celkovú úmrtnosť skúmanej populácie.
- ⌘ Vekovo špecifické úmrtnosti sa získajú z referenčnej populácie. Týmto spôsobom sa získa „očakávaný“ počet úmrtí v pozorovanej populácii očakávajúc, že vekovo špecifické úmrtnosti štandardnej populácie sa prepočítajú na sledovanú populáciu.
- ⌘ Najčastejšie sa uvádza pomer medzi pozorovanými a očakávanými úmrtiami, ktorý sa nazýva **index štandardizovanej úmrtnosti (Standardized Mortality Ratio SMR)**.

Nepriama štandardizácia

- ⌘ Je jedným z častých charakteristík populácie, na základe ktorej sa možno zhruba orientovať pri usudzovaní a porovnávaní zdravia medzi viacerými populáciami.
- ⌘ Je formou štandardizovania populácie a preto je možné priamo porovnávať LE medzi rôznymi populáciami bez štandardizácie.
- ⌘ Významnou výhodou je jednoduchá konštrukcia a možnosť populáciu charakterizovať jedným číslom. To je zároveň jej obmedzením. Za týmto číslom sa totiž skrýva história života, respektíve smrti generácie, ktorá tu už nie je. Preto úvahy, ktoré sa na základe tohto indikátoru zvyknú robiť, je potrebné konfrontovať s ďalšími ukazovateľmi, najmä zo štúdií stavu zdravia žijúcej populácie.
- ⌘ Napriek všetkému poskytuje veľmi užitočnú orientáciu a do veľkej miery dynamika jej vývoja charakterizuje zdravotné správa-nie populácie.

Očakávaná dĺžka života (LE)

- ⌘ Akákoľvek súhrnná miera môže prekryť faktory ktoré by mohli mať významné dôsledky pre zdravie verejnosti.
- ⌘ Napríklad pri štandardizácii na vek je možné prehliadnuť vekovo špecifické rozdiely v riziku v čase alebo mieste. Takáto situácia môže nastať vtedy keď výskyt nových prípadov ochorenia zdanlivo narastá v dôsledku efektu kohorty narodení (ľudia v mladšom veku môžu mať vyššie riziko ochorenia v porovnaní so staršími). Vekovo štandardizovaná miera môže skrývať tieto trendy.
- ⌘ Aj napriek tomuto riziku štandardizované miery poskytujú užitočné informácie, najmä v prípadoch zriedka-vých ochorení a pri širokej variabilite špecifických mier.

Obmedzenia štandardizácie

- ⌘ Štandardizácia priama – postup za použitia Excel a v {R}
- ⌘ Očakávaná dĺžka života
- ⌘ Výhody a interpretácia pre potreby porovnávania zdravotného stavu populácií medzi sebou

Súhrn

- ⌘ Vytvorte databázu úmrtností v BASE stiahnutím údajov z POPIN alebo z WHO. Vypočítajte v Excel a pomocou EPITOOLS sdr pre Slovenskú republiku, Českú republiku, Rakúsko a Maďarsko a interpretujte ich. Použite štandardnú populáciu SZO a priamu štandardizáciu.
- ⌘ Za použitia EPITOOLS vykonajte nepriamu štandardizáciu na tých istých údajoch.
- ⌘ Výsledky a obrázky postupu pošlite mgr. Dudákovej prostredníctvom MOODLE

Úlohy