

STATISTIKA 1

RNDr. K. Hrach, Ph.D.

Zápočet: 75% docházka na cvičení

+ odevzdání seminární práce (úkoly na PC)

Zkouška: písemná (bez kalkulačky, bez vzorců)

KONZULTACE

„Není hanba, že nevíš, ale že se neptáš.“ (*Turecké přísloví*)

STATISTIKA

Činnost vedoucí k získávání dat

Instituce zajišťující tuto činnost

Jakákoli shromažďovaná data

Údaje získané výpočtem z dat

Matematická teorie o chování dat

STATISTIKA

DESKRIPCE (popis)

ANALÝZA (modely, odhady, testy)

Základní pojmy

- STATISTICKÁ JEDNOTKA =
= na kom (čem) zjišťujeme
- STATISTICKÉ ŠETŘENÍ =
= jak zjišťujeme
- STATISTICKÁ VELIČINA =
= co zjišťujeme

STATISTICKÁ JEDNOTKA

s.j. = např. každý(á/é)

ptačí vejce (hmotnost v g?);

snůška (počet vajec?);

územní celek - obec; region;...; záměrně

selektovaný čtverec (% zalesnění?);

respondent (dotazníkové šetření)

...

STATISTICKÉ ŠETŘENÍ

■ ÚPLNÉ ⇔

informace od všech stat.jednotek
(od celé populace)

■ VÝBĚROVÉ ⇔

informace od vybraných stat. jednotek
(od „výběru“)

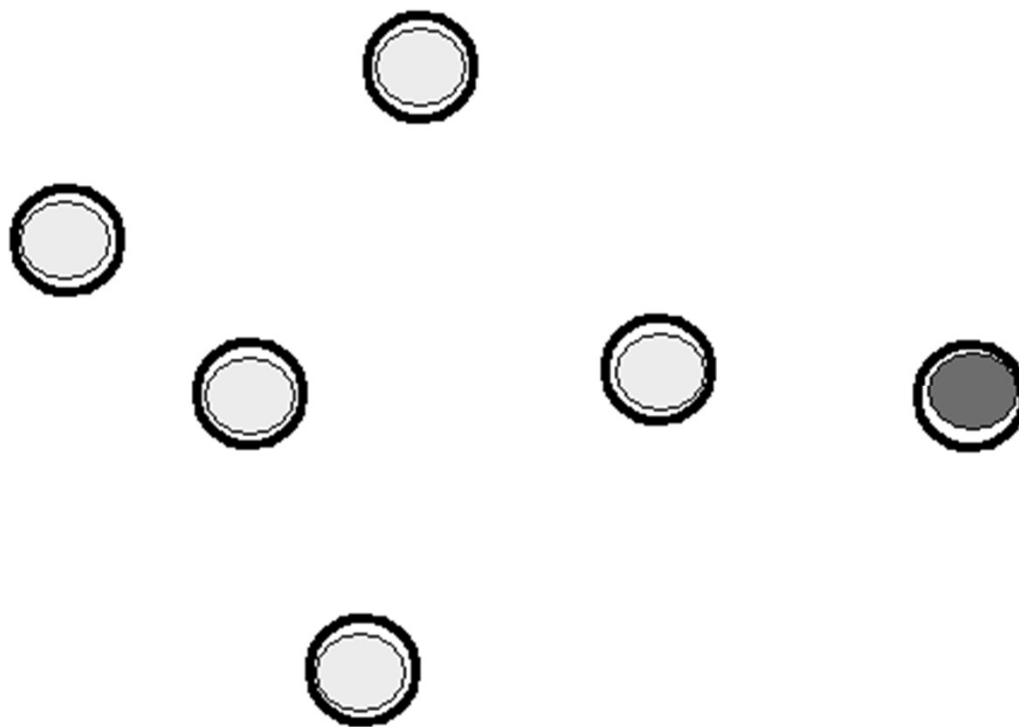
nevýhody versus výhody výběru?

* neúplnost informace

* rychlejší a levnější informace

VÝBĚROVÉ ŠETŘENÍ

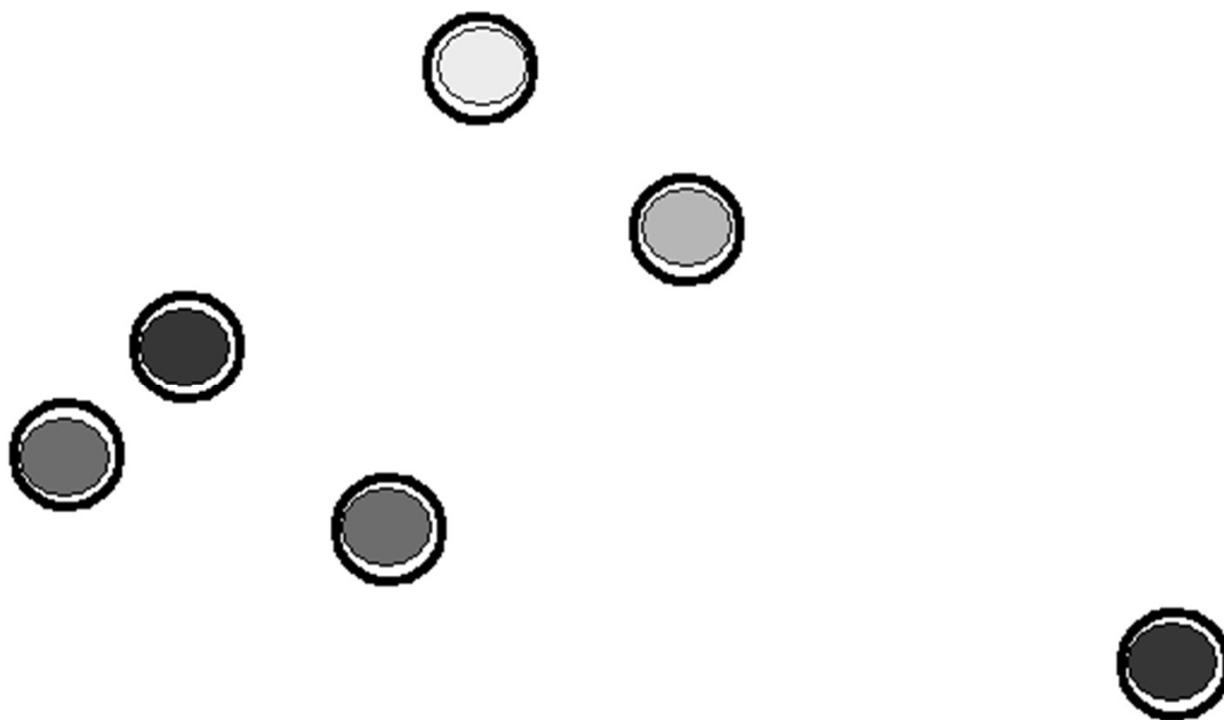
!!! NEREPREZENTATIVNOST !!!



VÝBĚROVÉ ŠETŘENÍ

!!! REPREZENTATIVNOST !!!

(zajištěna např. NÁHODNÝM výběrem)



VÝBĚROVÉ ŠETŘENÍ - výběr

- zcela náhodný
- systematický
- stratifikovaný ...

VÝBĚROVÉ ŠETŘENÍ - průběh

- vlastní měření ... s.j. je objekt
- anketa ... s.j. je subjekt (i dále):
- řízený rozhovor
- vyplnění dotazníku

Jak jinak získat data (např. k BP)?

JIŽ „HOTOVÁ“ (tj. sebraná), NAPŘ OD:

- ČSÚ (www stránky)
- Eurostatu (www stránky)
- úřadů státní (samo)správy
- firem (?) ...

Zpracování dat

a) „ručně“

b) pomocí SW

- MS Excel
- STATISTICA, SPSS, ...
- freeware (R-project)

DOTAZNÍK pro zaměstnance firmy

JMÉNO	
DOBA	
ÚKOL	
POHLAVÍ	
VZDĚLÁNÍ	
VĚK	
POBOČKA	
POČET DĚTÍ	

DOTAZNÍK pro zaměstnance firmy

JMÉNO	<i>identifikátor</i>
DOBA	<i>veličina (značena např. X)</i>
ÚKOL	<i>veličina (značena např. Q)</i>
POHLAVÍ	<i>...</i>
VZDĚLÁNÍ	<i>...</i>
VĚK	<i>...</i>
POBOČKA	<i>...</i>
POČET DĚTÍ	<i>veličina (značena např. Y)</i>

DOTAZNÍK pro zaměstnance firmy

	<i>možné hodnoty :</i>
JMÉNO	textový řetězec
DOBA	1,2,...(počet dní proškolení)
ÚKOL	ano / ne (splněn nový úkol?)
POHLAVÍ	m / z
VZDĚLÁNÍ	z, s, v (nejvyšší dosažené)
VĚK	v rocích
POBOČKA	a,b,c (1 ze 3 poboček firmy)
POČET DĚTÍ	0,1,2,...

DOTAZNÍK - příklad vyplnění

(1. statistická jednotka)

JMÉNO	Frank
DOBA	14
ÚKOL	ne
POHLAVÍ	m
VZDĚLÁNÍ	z
VĚK	30
POBOČKA	a
POČET DĚTÍ	1

DOTAZNÍK data (začátek)

Poradi	JMENO	DOBA	UKOL	POHL	VZDEL	VEK	POBOCKA	DETI
1	Frank	14	ne	m	z	30	a	1
2	Henry	29	ne	m	s	41	a	3
3	Tom	6	ne	m	v	32	b	0
4	Beth	25	ano	z	z	38	b	2
5	Susan	18	ano	z	s	44	b	2
6	Harry	4	ne	m	z	44	c	4
7	Paul	18	ne	m	s	39	c	4
8	Pete	12	ne	m	v	46	a	3
9	Diana	22	ano	z	z	45	a	2
10	Louise	6	ne	z	s	33	c	0
11	Fred	30	ano	m	z	32	a	1
12	Hank	11	ne	m	s	48	b	5
13	Steven	30	ano	m	v	42	b	3

DOTAZNÍK data (dokončení)

Poradi	JMENO	DOBA	UKOL	POHL	VZDEL	VEK	POBOCKA	DETI
14	Tod	5	ne	m	z	23	b	0
15	Take	20	ano	z	s	44	c	1
16	Sam	13	ne	m	z	34	c	2
17	Gail	9	ne	z	s	54	a	3
18	Thomas	32	ano	m	v	39	c	2
19	Theodore	24	ne	m	z	30	a	1
20	Charles	13	ano	m	s	39	a	3
21	Ellein	19	ne	z	z	51	b	0
22	Lori	4	ne	z	s	31	b	0
23	Ann	28	ano	z	v	38	c	1
24	Valerie	22	ano	z	z	35	c	2
25	Anke	8	ano	z	s	26	b	0

Značení dat (pozorování)

Např. veličina X – DOBA:

1. pozorování: $x_1=14$

2. pozorování: $x_2=29$

...

n . (poslední) pozorování: $x_n = x_{25} = 8$

n značí počet pozorování (rozsah souboru),
zde $n=25$

DOTAZNÍKY - značení a data

JMENO	DOBA (X)	UKOL (Q)	...	DETI (Y)
Frank	$x_1 = 14$	$q_1 = \text{ne}$...	$y_1 = 1$
Henry	$x_2 = 29$	$q_2 = \text{ne}$...	$y_2 = 3$
...
Anke	$x_{25} = 8$	$q_{25} = \text{ano}$...	$y_{25} = 0$

TYPY VELIČIN (X, Y, \dots)

(proměnných; znaků; angl. VARIABLE)

KATEGORIÁLNÍ

- alternativní (2 možnosti: 0-1)
např. POHL (m/z), UKOL (splněn: ano/ne)
- slovní – neuspořádané (nominální)
např. POBOCKA (a, b, c)
- slovní – uspořádané (ordinální)
např. VZDEL ($z < s < v$)
- číselné (diskrétní)
např. DETI (počet dětí)

TYPY VELIČIN (X, Y, \dots) (proměnných; znaků)

NEKATEGORIÁLNÍ

■ číselné (spojité)

např. DOBA (počet dní výcviku), VĚK
(v rocích), HMOTNOST (kg), PLAT (tis.Kč)...

v příkladech je údaj vlastně zaokrouhlen,
záleží na zvolené přesnosti;

lze převést na kategoriální typ (jak, jaký?)

DOTAZNÍKY – příklad zpracování (Y)

POČET DĚTÍ (Y)			kumul.	kumul.
y_i	n_i	p_i	n_i^*	p_i^*
0	6	0,24	6	0,24
1	5	0,20	11	0,44
2	6	0,24	17	0,68
3	5	0,20	22	0,88
4	2	0,08	24	0,96
5	1	0,04	25	1,00
suma	25	1,00	xxx	xxx

DOTAZNÍKY – příklad zpracování (Y)

POZOR NA PODOBNÉ ZNAČENÍ:

a) pro jednotlivá pozorování veličiny Y bylo

$$y_1=1, y_2=3, \dots, y_{25}=0 \quad (n=25)$$

b) pro kategorie veličiny Y bylo

$$y_1=0, y_2=1, \dots, y_6=5 \quad (K=6)$$

V praxi je rozdíl v použití jasný z kontextu.

ČETNOSTI

ABSOLUTNÍ

n_i ... počet výskytů i -té kategorie, $i=1 \dots K$

$$\sum n_i = n$$

RELATIVNÍ

p_i ... rel. výskyt i -té kategorie, $i=1 \dots K$

$$p_i = n_i / n \qquad \sum p_i = 1$$

$$p_i = (n_i / n) \cdot 100\% \qquad \sum p_i = 100 (\%)$$

Oba typy lze určit u každé kategoriální veličiny
(K =počet kategorií).

ČETNOSTI

Příklad 1.

Y - známky žáka. Popořadě:
3, 4, 2, 3, 2, 3, 3, 3.

Tabulka četností:

i	1	2	3	suma
y_i	2	3	4	xxx
n_i	2	5	1	8
p_i	0,250	0,625	0,125	1,000

ČETNOSTI

KUMULOVANÉ ABSOLUTNÍ

n_i^* ... počet výskytů do i-té kategorie včetně,

$$n_i^* = n_1 + \dots + n_i$$

KUMULOVANÉ RELATIVNÍ

p_i^* ... rel. výskyt do i-té kategorie včetně,

$$p_i^* = p_1 + \dots + p_i$$

$$p_i^* = n_i^* / n$$

Oba typy mají smysl jen u veličin
ordinálních či diskrétních.

ČETNOSTI

Příklad 1 - pokračování.

i	1	2	3	suma
y_i	2	3	4	xxx
n_i	2	5	1	8
p_i	0,250	0,625	0,125	1,000
n_i^*	2	7	8	xxx
p_i^*	0,250	0,875	1,000	xxx

ČETNOSTI

MODUS (skloňujeme: bez modu, ..., s modem)

Je (jsou) kategorie s největší četností
(samozřejmě nikoli kumulovanou).

Značen \hat{y} (se stříškou).

Lze určit u každé kategoriální veličiny.

Příklad 1 – pokračování:

$$\hat{y}=3$$

(druhá kategorie se vyskytla nejčastěji, a to
pětkrát; nejčastější známkou byla trojka)