



# Metody pro získávání dat, informací a poznatků v rámci veřejné strategie

## Příloha č. 5: Návod, jak použít šablonu v Excelu

Výstup v rámci aktivity KA1

Implementace metod a nástrojů pro strategické řízení a plánování

Aktivita 1.1 Metodická podpora pro strategické řízení a plánování

Nástroj č. 9/12: Metody pro získávání dat, informací a poznatků v rámci veřejné strategie

Název projektu: Podpora strategického řízení a plánování ve veřejné správě ČR  
s důrazem na rozšiřování metod, nástrojů, inovací, znalostí a zkušeností v praxi  
Registrační číslo CZ.03.4.74/0.0/0.0/15\_025/0016924



MINISTERSTVO  
PRO MÍSTNÍ  
ROZVOJ ČR

Název: Návod, jak použít šablonu v excelu – statistické výpočty. Příloha č. 5 nástroje  
Metody pro získávání dat, informací a poznatků v rámci veřejné strategie

autor: Milan Křápek

počet stran: 10



## Obsah

1.	Základní přehled fungování	3
2.	Práce s daty	3
3.	Podobnost	5
4.	Testy středních hodnot a rozptylů	6
5.	Regresní analýza	7
6.	Chi-square testy	8
	a. Chi-square test dobré shody	8
	b. Chi-square test nezávislosti	9



## 1. Základní přehled fungování

Připravená šablona obsahuje 6 listů. První list *Data* slouží k přípravě dat pro další výpočty, data připravená v tomto listu mohou být snadno použity v ostatních listech. Data která zde používáme mohou vycházet z vložené databáze, na posledním listu, nebo mohou být naše vlastní. Důležité pro správné fungování některých výpočtů je aby související údaje (například údaje o jedné obci v jednom období) byly na stejném řádku.

List *podobnost* dokáže na základě zvolených kritérií a jejich důležitosti určit srovnatelné obce. Následující listy jsou pak jednotlivé statistické analýzy. Nalezneme zde tedy listy, které počítají testy hypotéz o střední hodnotě a rozptylu (*Testy střední hodnoty a rozptylu*), list obsahující regresní analýzu (*Regresní analýza*) a nakonec list obsahující chi-square testy dobré shody a závislosti (*Chi-square testy*).

Poslední list *Data obce* obsahuje vloženou databázi, kterou pokud chcete můžete sami dle svého uvážení aktualizovat.

Předpokládaná činnost se šablonou spočívá v přípravě dat a jejich následujícím využití pro další výpočty. Není to však nutné všechny jednotlivé analýzy umožňují přímé vložení vlastních dat bez předchozí úpravy v listu *Data*. Zde je doporučování používat jiné názvy proměnných, než které máte nastavené v listu *Data*. Pokud použijete již existující název mohou být Vaše data přepsána údaji z listu *Data*.

## 2. Práce s daty

Práce s daty může spočívat v úpravě interní databáze na listu *Data obce*. Prakticky si tyto data můžete upravovat, jakkoliv chcete. Můžete přidávat nové proměnné, proměnné upravovat i mazat, přepisovat jednotlivé údaje i celé sloupce. Nic nebrání tomu abyste si zde vložili databázi obsahující i zcela jiné údaje než o obcích. Důležité pro správné fungování šablony ovšem je **neměnit název listu a v prvním sloupci mít identifikaci** podle které je každý řádek jednoznačně identifikovatelný a mít tabulku souvislou (bez prázdných popisků sloupců a identifikátorů řádků). Zde připravené údaje pak můžeme dále využívat v listu *Data*.



List *Data* obsahuje tabulku, kam můžete vkládat zkoumaná data. Ty sem dostanete buď kliknutím na nadpis sloupce a výběrem proměnné z rozevřacího seznamu (pozor jsou zde dvě stejné ikony, jedna je tento seznam a druhá otevírá filtry), nebo vlastním vložení. Jestliže zvolíte proměnnou ze seznamu, budou data načtena z databáze na listu *Data obce*, s tím, že první sloupec slouží jako identifikátor. Pokud nechcete použít tyto předpřipravené údaje, napište název sloupce sami tak aby nebyl totožný s některým v seznamu. To poznáte tak, že po napsání jiného názvu zůstane obsah sloupce beze změny. Poté již sami vyplníte údaje, které potřebujete.

Data můžete libovolně řadit a filtrovat pomocí připravených automatických filtrů, takto vyfiltrovaná data pak lze snadno použít k výpočtům na dalších listech.

ICO	Výdaje	poměr obyvatel 15-64	poměr obyvatel 65+
44992785	Výdaje zadlužení	0,639943	0,205934
845451	Daňové příjmy	0,65093	0,199047
75370	Nedaňové příjmy	0,640231	0,208651
262978	Kapitálové příjmy	0,633453	0,200613
299308	Přijaté transfery	0,633736	0,202799
244732	Běžné výdaje	0,633959	0,208969
268810	Kapitálové výdaje	0,613368	0,236559
274046	2196245187	0,63477	0,212917
81531	2204484137	0,633402	0,202847
782971	2496501793	0,621718	0,221236
	1799111799		

List obsahuje další dvě funkcionality, které usnadní práci s daty.

První spočívá v nastavení dat předem určených podobných obcí z listu *Podobnost*, to spustíme kliknutím na tlačítko *Použij data podobných obcí*. V takovém případě jsou původní identifikátory v tabulce vymazány a k nadpisům, které odpovídají databázi jsou vyhledány odpovídající hodnoty. Sloupce s jinými názvy zůstanou beze změny.

Druhou pak použijete v případě, že jste ručně upravili sloupec s identifikátory (první sloupec dat) a chcete, aby šablona znovu doplnila ostatní sloupce. Funkcionalitu spustíme tlačítkem *Doplň ostatní sloupce dle IČO*.



### 3. Podobnost

Na listu podobnost musíme nejdříve nastavit základní informace. Určíme tedy *IČO*, po jeho vyplnění je automaticky ověřeno dle údajů v databázi ares a systém doplní název obce a adresu městského/obecního úřadu.

Potom určíme počet nejpodobnějších obcí, které chceme zjistit. Nakonec nastavíme podmínky pro určování podobnosti. Podobnost je určována z interní databáze, takže je nutné zvolit jednu z navržených proměnných, a navíc ty které jsou číselné. U každé proměnné pak určíme váhu. Ta udává důležitost této proměnné ve srovnání s ostatními.

Název obce	IC	rozdílnost
Brno (okres Brno-město)	44992785	4,595721
Ostrava (okres Ostrava-město)	845451	5,198701
Plzeň (okres Plzeň-město)	75370	5,464275
Liberec (okres Liberec)	262978	5,649957
Olomouc (okres Olomouc)	299308	5,678478
České Budějovice (okres České Budějovice)	244732	5,697744
Hradec Králové (okres Hradec Králové)	268810	5,707883
Pardubice (okres Pardubice)	274046	5,733907

Celá metoda funguje tak, že pro všechny obce v databázi program určí rozdíly proměnných se zkoumanou obcí. Každou takto získanou diferenci znormuje aby nezávisely na absolutních rozdílech a elativních a poté výsledek vynásobí váhou, čímž proměnná získá určenou důležitost. Pro důležitosti tak nejsou podstatně absolutní rozdíly ale poměry vah. Například pokud určíme výhu 1 a pět, tak proměnná, která má váhu 5 je pětkrát důležitější než proměnná s váhou 1. Použitý rozsah čísel není nijak ovlivněn a je tedy zcela na Vás.

Výpočet pak spustíme tlačítkem *Najít podobné obce*. Výsledky se pak zobrazí v tabulce vpravo.

## 4. Testy středních hodnot a rozptylů

Tento list umí vyhodnotit několik základních testů střední hodnoty a rozptylu popsaných v metodice.

Jedná se buď o jednovýběrové testy, kdy je použit jen první sloupec s daty. Párové testy a dvouvýběrové testy, tyto pak potřebují dva sloupce dat, s tím že u párového je nutné, aby sloupce měly stejný počet vyplněných řádků, u dvouvýběrového to nutné není. Jakou variantu chceme počítat určíme výběrem vlevo. Také musíme určit hladinu významnosti Alfa.

**Dvouvýběrový t-test**

Jednovýběrový test střední hodnoty a rozptylu  
 Párový t-test  
 Dvouvýběrový t-test F-test

Alfa: 0,05

Aktualizuj data

Soubor dat X	Soubor dat Y
poměr obyvatel 15-64	poměr obyvatel 65+
0,639942629	0,205933577
0,650929692	0,199046789
0,640230572	0,208651075
0,633453014	0,200612763
0,633735563	0,202799359
0,633958772	0,208968877
0,613368269	0,236559488
0,63476976	0,212916814
0,633402186	0,202846631
0,624718322	0,224235636

Statistiky	Soubor X	Soubor Y
Počet hodnot	15	15
Aritmetický průměr	0,637291031	0,209680516
Výběrová směrodatná odchylka	0,011028062	0,014301766

Testové kritérium	0,324789699
Kritický obor	$(-\infty; -2,0484) \cup (2,0484; \infty)$
P-hodnota	2,94125E-36

Na hladině významnosti 5% můžeme potvrdit, že střední hodnoty souborů se liší

Data můžeme načíst z listu *Data* výběrem názvu sloupce z rozevřacího seznamu, nebo sami vložit.

Pokud použijeme rozevřací seznam je sloupec, u kterého jsme nastavili nadpis rovnou vyplněn s tím, že budou zobrazeny jen ty údaje, které jsou právě viditelné. Viditelnost řádků dat můžeme určit filtrem na listu *Data*. Data nejsou načítána automaticky, pokud bychom tedy později upravili údaje v listu *Data*, a to buď změnu, nebo nový filtr, změny se neprojeví v tomto listu, nehrozí tedy, že výsledky ztratíme. Pokud chceme, aby se provedené změny propsaly do tohoto listu musíme použít tlačítko *Aktualizuj data*.

V pravé části pak uvidíme výsledky příslušných testů, dle nastavení.



## 5. Regresní analýza

V případě regresní analýzy vycházíme z lineárního modelu s jednou vysvětlovanou proměnnou a několika vysvětlujícími. Jejich počet určíme vyplněním přirozeného čísla do příslušné buňky

Dle tohoto nastavení se zobrazí potřebný počet sloupců.

Počet vysvětlujících proměnných: 2

Spust regresní analýzu

**Regresní analýza**

VÝSLEDEK

Regresní statistika	
Násobné R	0,079545
Hodnota spolehlivosti R	0,006327
Nastavená hodnota spolehlivosti	-0,02794
Chyba stř. hodnoty	2631,229
Pozorování	31

**ANOVA**

	Rozdíl	SS	MS	F	znamnost F
Regrese	1	1277491	1278491	0,184663	0,670572
Rezidua	29	2,01E+08	6923368		
Celkem	30	2,02E+08			

	Koeficienty	stř. hod.	t Stat	Hodnota P	Dolní 95%	Horní 95%
Hranice	1795,424	523,7285	3,428158	0,00184	724,2789	2866,569
Soubor X 1	-3,7E-06	8,56E-06	-0,42972	0,670572	-2,1E-05	1,38E-05

Aktualizuj data

číslo položky	Y	X1	X2
1	0,172307692	34712220,39	
2	0,193548387	103075262,4	
3	0,178046934	10914376,35	
4	0,221627409	1560289,16	
5	0,165656566	43891914,62	
6	0,150197628	61711682,69	
7	0,163522013	6526251,84	
8	0,118181818	15176058,07	
9	0,219281664	29407844,87	
10	0,207706422	17475233,85	
11	0,207920792	11663754,96	
12	0,177356021	5144034,91	
13	0,184381779	13449111,38	
14	0,219672131	7374927,6	
15	0,152046784	3478990,82	
16	0,189	20164120,47	
17	0,163043478	1879909,63	
18	0,203208556	5193375,41	
19	0,12	7569552,49	
20	0,157814871	307083687,8	

Data můžeme načíst stejně jako při testování hypotéz na předchozím listu, tedy výběrem z listu *Data* výběrem názvu sloupce z rozevřacího seznamu, nebo sami vložit. Zde je ovšem proti předchozímu listu nezbytné, aby každý řádek obsahoval související údaje a všechny sloupce obsahovaly stejný počet vyplněných hodnot.

Kliknutím na tlačítko *Spust regresní analýzu* získáme výsledky v levé části okna.







následující interval začíná tam, kde předcházející končí. První interval je vždy počítán od nuly bez ohledu na skutečnou hodnotu v intervalu zapsanou. Poslední je vždy do n.

**Poissonovo rozdělení** opět umožňuje jen číselné hodnoty s tím, že dolní hodnota by měla být nula, horní je nekonečno, a tedy jak v případě intervalů, tak v případě obměn bude poslední hodnota počítána jako interval končící v nekonečnu.

**Hypergeometrické rozdělení** umožňuje obměny i intervaly, ale v případě obměn je nutné, aby se jednalo o čísla. Hodnoty obměn musí pokrývat všechny možné obměny zvoleného binomického rozdělení. Tedy od 0 do n. V případě intervalů platí, že na sebe musí přímo navazovat, tedy vždy následující interval začíná tam, kde předcházející končí. První interval je vždy počítán od nuly bez ohledu na skutečnou hodnotu v intervalu zapsanou. Poslední je vždy do n.

**Normální rozdělení** předpokládá číselnou proměnnou ve formě intervalů a krajní intervaly jsou upraveny jako nekonečné.

Výsledky testu jsou počítány automaticky a najdeme je v dolní části okna.

## b. Chi-square test nezávislosti

U tohoto testu je nutné určit si dvě proměnné dle kterých dojde ke třídění. V šabloně nastavíme hodnotu Alfa, počet řádků a počet sloupců. Poté vyplníme popisky obměn, a hlavně jejich absolutní četnosti. Po vyplnění je již automaticky vše dopočítáno a pod nastavením se zobrazí výsledek testu.

Chi-square test závislosti

alfa	0,05
počet skupin první proměnné	4
počet skupin druhé proměnné	3

p-hodnota chi-square testu 0,00000

Sledované proměnné jsou statisticky závislé na hladině významnosti 5%.

Empirické četnosti

Obměny	Základní	Pověřený obecní úřad	Rozšířená působnost	
0-500	3414	0	0	3414
501-3000	2317	61	1	2379
3001-30000	126	122	169	417
více než 30000	0	0	36	36
	5857	183	206	6246

Teoretické četnosti

Obměny	Základní	Pověřený obecní úřad	Rozšířená působnost	
0-500	3201,377	100,0259	112,5975	3414
501-3000	2230,836	69,70173	78,46206	2379
3001-30000	391,0293	12,21758	13,75312	417
více než 30000	33,75793	1,054755	1,18732	36
	5857	183	206	6246



**Červeně** označené jsou hodnoty, jejichž teoretické četnosti nespĺňují podmínku, pokud se takové objeví, naznačuje to možnou nepřesnost výsledků. To můžeme řešit tak, že buď zvýšíme absolutní četnosti, tedy velikost vzorku, nebo sloučíme některé obměny.

V tomto případě bychom mohli sloučit dva poslední řádky a test by byl průkazný.